# Auditory perceptual learning via decoded EEG neurofeedback: a novel paradigm

Alex Brandmeyer

# Auditory perceptual learning via decoded EEG neurofeedback:
## a novel paradigm

**Proefschrift**

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 19 maart 2014
om 12.30 uur precies

door

**Alex Brandmeyer**

geboren op 24 november 1980 te La Jolla, California, USA

**Promotoren**:
Prof. dr. Peter Desain
Prof. dr. James M. McQueen

**Copromotor**:
Dr. Makiko Sadakata

**Manuscriptcommissie:**
Prof. dr. John van Opstal
Prof. dr. Nick Ramsey (UMCU)
Dr. Marcel van Gerven

# Contents

# 1

# Introduction

*Imagine you are on the edge of a lake and a friend challenges you to play a game. The game is this: your friend digs two narrow channels up from the side of the lake. Each is a few feet long and a few inches wide, and they are spaced a few feet apart. Halfway up each one, your friend stretches a handkerchief and fastens it to the sides of the channel. As waves reach the side of the lake they travel up the channels and cause the handkerchiefs to go into motion. You are allowed to look only at the handkerchiefs and from their motions to answer a series of questions: How many boats are there on the lake and where are they? Which is the most powerful one? Which one is closer? Is the wind blowing? Has any large object been dropped suddenly into the lake? Solving this problem seems impossible, but it is a strict analogy to the problem faced by our auditory systems.*

– Albert S. Bregman, *Auditory Scene Analysis*

*When I say 'love', the sound comes out of my mouth, and it hits the other person's ear, travels through this byzantine conduit in their brain, through their memories of love or lack of love, and they register what I'm saying and they say 'yes,' they understand, but how do I know they understand because words are inert, they're just symbols...*

– Kim Krizan, from the film *Waking Life*

*If we spoke a different language, we would perceive a somewhat different world... The limits of my language are the limits of my world.*

– attributed to Ludwig Wittgenstein

Our experiences of sound are so much more than just sound: the notes, rhythms and instrumentation of our favorite songs, the familiar voices of colleagues, friends and loved ones during conversations, or the powerful words of a moving speech. All of these sounds are deeply imbued with meanings we have learned and that are prescribed by the context in which we experience them. Yet we perceive these sounds effortlessly and transparently, unaware of the complex neurophysiological mechanisms that underlie our experiences.

This dissertation explores these abilities and the underlying neurophysiology, and presents the development of an approach that attempts to monitor these perceptual processes in real-time, with the aim of realizing technologies that can aid us in improving our abilities to perceive new and unfamiliar sounds. Specifically, it presents the development a brain-computer interface paradigm based on electroencephalographic (EEG) measurements of auditory evoked potentials containing the mismatch negativity response. This paradigm provides users with neurofeedback based on the decoding of single-trial evoked potentials, the goal of which is to enhance the perception of domain relevant sound contrasts, such as foreign language phonemes or pitched tones. The concepts and research findings upon which the present work is based are introduced here in this first chapter.

Afterwards, the results of a series of preliminary analyses using data collected from native and non-native speakers of English during speech perception are presented. The goal of these analyses was to determine the feasibility of a neurofeedback approach based on single-trial pattern classification of EEG data measured during the perception of the types of sound contrasts envisioned for use in our neurofeedback paradigm. Subsequently, the results of an initial experiment are presented in which the effects of this neurofeedback paradigm on behavioral discrimination abilities and brain responses are explored. Finally, a methodological framework for real-time tracking of the brain responses underlying perceptual discrimination is presented.

## Auditory Perception, Language and Music

The introductory quote from Al Bregman's seminal book on auditory perception highlights the amazing task with which our ears are tasked. Using just the vibrations of air against two flaps of skin inherited from our fish ancestors, evolution has produced a complex and elegant system of bones, organs and nerves that enable us to perceive sound.

The quote also points out a crucial difference between sensation (the specific way in which patterns of air vibrations stimulate our central nervous system) and perception (our rich and meaningful experience of sounds). Sound is sensed by the ear using tiny hairs that move in currents of fluid inside the cochlea, part of the inner ear. The movement of these hairs opens ion gates embedded in the membrane of the cell to which they are attached, triggering a receptor potential. This leads to the firing of the auditory nerve cells adjacent to them, which sends a cascade of signals through progressively higher levels of the auditory nervous system that extract information about the sound, such as its location (Zwiers et al., 2004), and whether it is moving. These processes form representations of auditory 'streams' (Bregman, 1990) or 'objects' (Griffiths and Warren, 2004). A mixture of information from the many sound sources around us contained in the complex activity patterns of the peripheral auditory nerve is processed by the ascending stages of the auditory pathway, and is eventually transformed into working knowledge of the type, number and location of the sound-generating objects around us.

The ability to perceive individual objects and events in the world within mixtures of sounds lies at the heart of some of our most fundamental experiences. Take for example our ability to communicate using language, a predominantly auditory and motor skill that is unique to humans. The capacity to produce and perceive meaningful sequences of sounds sets us apart from other animals and is a defining feature of human society. Closely related to language is music, whose evolutionary significance has been tied directly to the development of our linguistic capabilities (Patel, 2010), and which exists in every known culture. A comprehensive understanding of these unique abilities is one of the principal tasks of the cognitive sciences.

Research into the neural underpinnings of these abilities has revealed a complex orchestration of brain regions and networks that carry out the component processes. In language, tasks such as phoneme and word identification and syntactical processing are known to involve distinct brain networks and take place at specific points in time relative to one another (Friederici, 2002). Similarly, the perception of key musical dimensions such as rhythm, melody and harmony have been shown to involve distinct, parallel brain networks operating together that serve the formation of our holistic experiences of music (Koelsch and Siebel, 2005).

Evidence from studies involving patients with focal brain lesions has shown that damage to these specific brain networks can disrupt individual component processes while leaving others intact. This leads to disorders such as amusia (Peretz and Coltheart, 2003), in which individuals who previously enjoyed music are no longer able to perceive its melodic structure, or specific forms of aphasia, in which semantic or syntactic aspects of speech perception (Ogar et al., 2011) and production (Gorno-Tempini et al., 2011) become impaired. These ailments highlight both the complexity and the fragility of our ability to perceive language and music, and also provide insight into the cognitive mechanisms underlying our ability to learn in these domains.

## Perceptual Learning

We are not born into the world with fully developed linguistic and musical abilities. Babies come into this world without the ability to speak or identify meaningful sounds in their parents' language(s). Through exposure to these sounds in their environment, children already begin to show signs that they have acquired knowledge of their native language by six months of age (Kuhl et al., 1992), and begin to start speaking shortly thereafter. By adulthood, individuals show a strong perceptual bias to the sounds in their own language, at the expense of being able to perceive meaningful differences in a foreign language (Goto, 1971; Iverson et al., 2003).

Studies investigating language learning often focus on the perception of phonemes, which are the fundamental sound units of language. Humans are capable of producing and perceiving a wide variety of phonemes, but each individual language employs only a small subset of the possible phonemes. Which particular subset of phonemes is used by a given language, in combination with other factors such as timing patterns, gives each language its unique character and sound. Differences in phonetic makeup represent just one of many factors that can make learning second languages difficult.

Crucially, phonemes are perceived categorically. Take for example the English sounds /p/ and /b/, as in the words "pet" and "bet". Both are described as labial stop conso-

nants by the International Phonetic Alphabet (IPA), a system for classifying speech sounds on the basis of how they are articulated using the vocal tract (International Phonetic Association, 1999). Such consonants are realized though a closure of the lips during the initial movements of the vocal tract, leading to increased air pressure inside the mouth. When the lips open, the rushing of air out of the mouth leads to a plosive burst of sound, a key acoustic feature of these sounds. The difference between /p/ and /b/ is the presence of voicing (vibrations of the vocal folds): /p/ is unvoiced while /b/ is voiced. In turn, the two sounds are characterized by differences in voice-onset time (VOT), an acoustic feature that is an important determiner of which category of sound a listener will perceive (Keating et al., 1981). When presented with a continuum of sounds with varying VOT, responses typically follow a sigmoidal, or S-shaped pattern: responses sharply fall into one or the other category, with the point where the dominant response category changes referred to as the 'category boundary'.

Native speakers of English have learned over time that VOT cues provide the most important information needed to determine whether someone is talking about a pet or a bet. They are able to do this across a huge range of variability in the constituent sounds, realized by different speakers (children, adults, males and females) at different speaking rates (slow or fast, which affects VOT) and in different environments (quiet, noisy, indoors, outdoors). The ability to ignore this variability and instead ascribe a speech sound to one of a discrete number of classes (i.e. /p/ or /b/) is the hallmark of categorical perception, and reflects many years of experience learning our languages.

In contrast, our ability to perceive different categories of phonemes in a foreign language is often quite poor relative to native speakers, due to our lack of experience with those categories and the way in which they overlap with those of our own native language (Cutler, 2012). A well-known example of this phenomenon is the difficulty faced by speakers of Japanese when learning to distinguish the English language sounds /r/ and /l/ (Goto, 1971; Miyawaki et al., 1975). These sounds are articulated using distinct configurations of the vocal tract that lead to a difference in relative heights of the first and second formants. As Japanese only makes use of one sound that is somewhere between the two English sounds, native Japanese speakers untrained in English often incorrectly categorize these sounds, which can lead to misunderstandings at the level of words. Acquiring these new categories as an adult is one of the most challenging aspects of learning to speak and understand a new language, regardless of which language is being learnt.

Similar effects of learning and experience on music perception have also been observed. At an early age (approximately 6 months), children do not show preferences for culturally specific rhythms or scales (Lynch et al., 1990). By the age of about 10, children have difficulty in recognizing mistakes such as out of tune or mistimed notes in unfamiliar musical systems (Lynch and Eilers, 1991; Stalinski and Schellenberg, 2012), but can easily produce and perceive music that is familiar to them. Additionally, extensive musical training leads to relative enhancements of auditory perception. Musicians typically show enhanced discrimination of the principal musical dimensions of pitch (Micheyl et al., 2006) and rhythm (Rammsayer and Altenmüller, 2006) relative to untrained listeners. Interestingly, there is converging evidence that musical training also facilitates improved linguistic abilities in both first and second languages (Patel and Iversen, 2007; Asaridou and McQueen, 2012).

As with language, specific features of music such as pitch intervals (Siegel and Siegel, 1977) and note intervals (Desain and Honing, 2003) are perceived categorically. Which specific categories an individual perceives is strongly influenced by their culture, due to the

cultural specificity of different musical systems. While a Western listener will be familiar with diatonic note intervals, they might have difficulty in distinguishing the note intervals used in an unfamiliar system, such as the Javanese pelog scale used in Gamelan music (Krumhansl, 2000). Other studies have shown an influence of culture on the production of different categories of musical note intervals (Sadakata et al., 2004).

All of these findings suggest that our perception becomes tuned to the specific linguistic and musical environments that we inhabit. We have in fact become experts in these domains. Through further training in new languages or musical traditions our ability to acquire additional expertise continues into adulthood. Such expertise is a direct result of a domain-general phenomenon known as perceptual learning, a process in which our perceptual skills improve through exposure and practice. Such learning leads to changes in our behavior (e.g. the ability to recognize different categories of sounds) and in the way our brains respond to and process perceptual information. Through plastic changes in the patterns of connectivity between neurons in different regions of the brain, our perception of specific types of stimuli is refined through increased familiarity (Li et al., 2010). Using neuroimaging techniques such as EEG and functional magnetic resonance imaging (fMRI), it is possible to measure changes in the brain responses of participants during the course of perceptual learning (Reinke et al., 2003; Tremblay et al., 1998; Menning et al., 2000), as well as individual differences in these responses that correspond to differences in these individuals' perceptual skills (e.g. the ability to understand a specific language). In the case of music, it has been argued that the musician's mind should be taken as a 'model of neuroplasticity' (Münte et al., 2002) due to the extensive training effects that have been observed in neuroanatomical (Schlaug et al., 1995; Keenan et al., 2001) and neuroimaging (Schmithorst and Holland, 2003) studies.

As our understanding of the neural basis of perceptual learning has increased, so has interest in applying this knowledge in the context of education (Tallal and Gaab, 2006; Bishop, 2007; Meltzoff et al., 2009) and in the development of novel teaching methods that are based in part on the results of neuroimaging research (Goswami, 2006; Varma et al., 2008). Online measurements of specific types of brain activities associated with neuroplasticity and perceptual learning would provide both educators and students themselves important information about the mental state of the student that could be used to further enhance the learning environment. This thesis aims to identify these specific forms of brain activity, and to develop methods for monitoring them in real-time.

## Auditory Evoked Potentials and the Mismatch Negativity

A tool that has been invaluable in the study of auditory perceptual learning (and auditory perception in general) is the auditory evoked potential (AEP), which is a type of event-related potential (ERP). AEPs are measured by recording the brain's time-locked responses to sound presentations using EEG, and have been used to study the way in which the brain processes sound. Using the high temporal resolution of EEG, it is possible to investigate the sequence of neural activations at distinct locations along the auditory pathway triggered by the presentation of sounds. These activations, corresponding to the synchronous firing of populations containing thousands of neurons, can be measured as small voltages ($< 10$ $\mu V$) using electrodes placed at different locations on the scalp. They appear as a waveform with a series of positive and negative peaks when the time-locked average of the data is

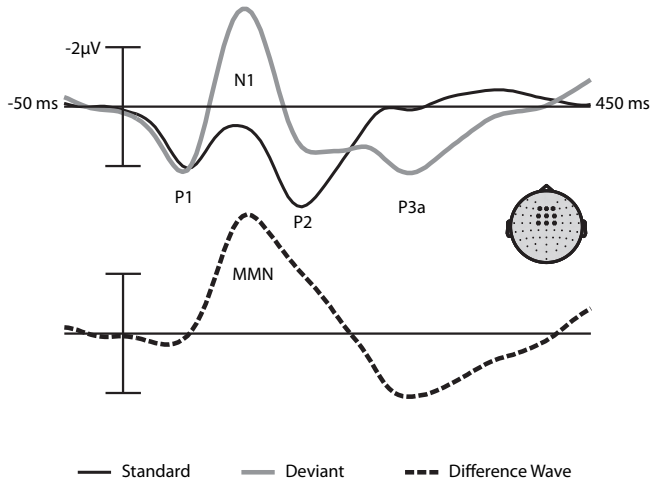


Figure 1.1: *Typical auditory evoked potentials (AEPs) elicited in a mismatch negativity paradigm. The average AEPs obtained for the high-probability standard stimulus and the low-probability deviant stimulus, as well as the deviant - standard difference wave are shown. The nine fronto-central electrode locations used to obtain the AEPs are indicated in the scalp map. These ERPs were obtained using sine wave tones with fundamental frequencies of 500 Hz (standard) and 600 Hz (deviant). The waveforms obtained for both the standard and deviant stimuli clearly show the P1-N1-P2 complex. In addition, the deviant waveform shows an enhanced negativity in the N1 time window (the mismatch negativity, or MMN) as well as an additional component known as the P3a, which is observed when a salient stimulus contrast is employed, and which reflects the engagement of attention triggered by the deviant stimulus.*

plotted. Analyzing the characteristics of these waveforms, such as the timing and amplitude of specific peaks (typically referred to as components), provides insight into the function of the populations at different locations within the central nervous system.

AEPs (and ERPs in general) are widely used in the study of cognitive processes, such as those underlying language (Friederici, 2002) and music (Koelsch and Siebel, 2005), as well as in medical and clinical settings (Bukard et al., 2007), where abnormalities in the AEP can be used in the diagnosis of specific disorders (e.g. damage to the auditory pathways) and conditions (e.g. schizophrenia). Typically, these studies use measurements of one or more components in the averaged waveforms obtained at different electrode recording sites. For instance the 'P1-N1-P2' complex is a series of positive (P) and negative (N) peaks in the AEP occurring between 50-200 ms following the onset of a sound, and is generated in the auditory cortex (Martin et al., 2007). When measured using EEG, these components typically have peak amplitudes at fronto-central electrode locations. This is due to the manner in which activity in bilateral auditory cortices projects to the scalp. Changes in the amplitude and timing of these components have been attributed to effects of perceptual learning (Tremblay et al., 2001).

One component of the AEP widely used to study the effects of perceptual learning is the mismatch negativity (MMN). An example of an AEP containing the MMN response can be seen in Figure 1.1. First reported by Näätänen in the late 1970s (Näätänen et al.,

1978), the MMN is elicited following unexpected changes in a repetitive sequence of sounds. It is typically measured as an increased fronto-central negativity between 100-300 ms following the onset of a so-called 'deviant' stimulus randomly embedded within repeating trains of high probability 'standard' stimuli (Näätänen et al., 2007). The MMN is a pre-attentive response, and can be measured in a passive listening paradigm in which participants perform a non-auditory task such as watching a silent film. Crucially, the MMN has been linked to our abilities to discriminate sounds from one another. Two sounds that can be reliably discriminated will elicit an MMN when presented in the oddball sequences described above. Conversely, the MMN is typically absent or reduced when using pairs of sounds that we are not able to discriminate (Näätänen, 1995).

This aspect of the MMN response has led to its widespread use in research investigating the effects of learning and expertise on auditory perception. Specifically, many language studies looking at native/non-native differences in phoneme perception have shown clear differences in the average MMN responses measured in groups of native and non-native speakers (Näätänen, 2001). Typically, non-native speakers will show an absent or reduced MMN response to a foreign phonetic contrast, while native speakers show a clear MMN. For example, Näätänen and colleagues conducted an experiment using native speakers of Finnish and Estonian along with a set of three vowel stimuli representing similar but distinct phonetic categories in Estonian (Näätänen et al., 1997). The key manipulation here was that while Estonians would clearly perceive these three vowels as belonging to different categories, only two of these vowels corresponded to phonetic categories in the Finnish language, while the third fell somewhere between the other two categories. This difference in categorical perception was reflected in the MMN responses measured in the two groups of participants: while the Estonian group showed clear MMN responses to all three vowels, the Finnish group's response to the third, ambiguous category was reduced relative to that of the Estonian group. While this study was one of the first to demonstrate such perceptual learning effects on MMN responses to speech stimuli, many have since demonstrated similar effects in other language pairs (Sharma and Dorman, 2000; Winkler et al., 1999), emphasizing the common neurophysiological mechanisms underlying language learning.

Other studies have used the MMN to explore the effects of musical expertise on perception. Musicians will typically show large amplitude and/or lower latency MMN responses to musically relevant sound contrasts as compared to musically untrained listeners (Münte et al., 2002). For example, stringed instrument players are typically required to discriminate very small differences in the pitch of two sounds in order to properly tune their instruments. Using sequences of musical chords that occasionally contained a slightly out-of-tune note, it was shown that this type of training in pitch discrimination leads to relatively large MMN responses as compared to an absent response in musical novices (Koelsch et al., 1999). Similar effects have been shown for more complex musical features such as melody contour (Fujioka et al., 2004), chord structures (Brattico et al., 2009) and note numerosity (van Zuijen et al., 2005).

The effects of perceptual learning on the MMN response have also been demonstrated by longitudinal training studies showing progressive enhancements of the responses measured using speech (Tremblay et al., 1998) and music stimuli (Menning et al., 2000). These changes are accompanied by improved accuracy in an individual's ability to behaviorally discriminate the targeted sounds. Interestingly, Tremblay and colleagues reported that, for about half of the participants in their language training study, an enhancement of the

MMN response to the trained speech contrasts preceded a corresponding change in behavioral discrimination. This suggests that plastic changes in auditory cortex associated with the enhancement of the MMN response represent a distinct stage of perceptual learning that is prior to improvements in behavioral discrimination.

This strong relationship between the MMN and auditory perceptual learning suggest that MMN measurements could serve as diagnostic markers of individual perceptual abilities in educational contexts. For example, instructors could adapt the difficulty of a given learning task on the basis of the MMN response, or even provide feedback to their students on changes in the measured MMN responses over time. However, the time typically required to measure the MMN represents an obstacle to its widespread use in educational settings. This is not unique to the MMN response, but merely reflects the nature of event-related potentials and other neuroimaging data analysis methods based on averaging large amounts of data. Recently however, a different family of methods known as multivariate pattern classification analysis is being used to analyze high-dimensional neuroimaging data at the level of the single-trial as well as to develop applications involving real-time neuroimaging.

## Machine Learning, Neuroimaging and Brain-Computer Interfaces

Pattern classification analysis is one of a number of different techniques developed in the field of Machine Learning. The goal of these approaches is to learn to recognize patterns in a specific type of data corresponding to two or more potential classes. Classic examples of machine learning applications include optical character recognition (Mori et al., 1999) and automatic spam email checkers (Cormack, 2007). Pattern classification analyses are usually based on one of a number of multivariate statistical methods, such as support vector machines, logistic regression or linear discriminant analysis (Bishop, 2009). Such methods can be applied to datasets with an arbitrary number of dimensions. One of the results of such an analysis is a statistical model of the data that can be used to make predictions about new data examples. These models are typically referred to as 'classifiers'.

The use of techniques from the field of machine learning in conjunction with neuroimaging data allows researchers to make predictions about specific forms of brain activity, such as those related to perception and learning, in real-time using only single measurements (Haynes and Rees, 2006; Blankertz et al., 2011). This process, often referred to as decoding, lies at the heart of so-called 'brain-reading' approaches and has been used to investigate cognitive phenomena including music (Schaefer et al., 2011) and language (Formisano et al., 2008) perception, as well as visual perception (Kay et al., 2008), attention (Kamitani and Tong, 2005) and consciousness (Haynes and Rees, 2005a,b).

It has also led to the proposal of a number of potential brain-reading applications, such as the ability to detect awareness in patients in vegetative states (Owen, 2006; Haynes and Rees, 2006) and novel forms of lie detection (Davatzikos et al., 2005). With respect to clinical and medical applications of the MMN response, it has recently been reported that decoding analyses of MMN responses measured in comatose patients can even be used to predict their chances of survival (Tzovara et al., 2013).

Another key application of pattern classification analysis techniques to neuroimaging

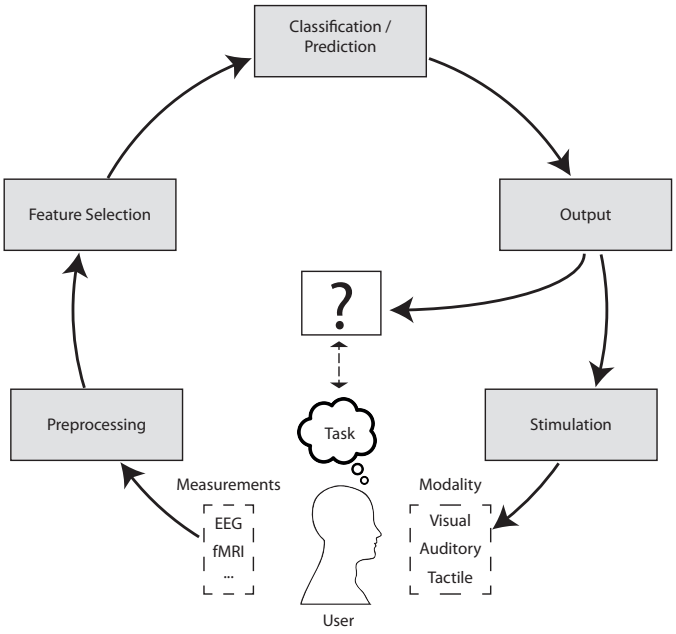


Figure 1.2: *A schematic of the BCI Loop, adapted from van Gerven et al. (2009). BCI design is typically thought of as a loop centered around the BCI user. Through one or more types of neuroimaging measurement, data are obtained while a user performs a specific type of mental task, such as attending to particular stimuli that have been presented in the visual or auditory modalities. These data are then preprocessed (e.g. filtering, re-referencing, artifact removal), and task-appropriate features in the data are selected, such as specific sets of voxels measured using fMRI, or electrode locations and time windows containing specific ERP components. A classifier is then applied to these features of the data, the output of which is a prediction regarding the mental state of the user (e.g. attending to item 'A' or item 'B'). This prediction is used to update the state of the BCI and stimulation parameters, leading to the next iteration of the loop.*

data is the development of brain-computer interfaces (BCIs, also referred to as brain-machine interfaces; for a recent review, see van Gerven et al., 2009). A BCI is a system that allows a user to operate a specific interface, such as a communication device (Farwell and Donchin, 1988; Schreuder et al., 2010; van der Waal et al., 2012), menu system (Schalk et al., 2004), graphical-user interface (Li et al., 2010), prosthesis device (Hochberg et al., 2013) or wheelchair (Leeb et al., 2007) using measurements of brain activity alone. These systems are principally designed for individuals who can no longer use other types of interfaces, such as patients with neuromuscular disorders or forms of paralysis. However, there has also been an increasing interest in the use of BCIs by healthy users, in contexts such as gaming (Coyle et al., 2013). The design of these systems is often conceptualized as a loop, a schematic of which is presented in Figure 1.2. This loop centers on the interplay between the mental state of the user and the BCI system's ability to make predictions about that mental state on the basis of brain data.

Users of a BCI typically perform a specific mental task, such as attending to a particular item on a screen (i.e. systems based on the P300 response, see for example Sellers and Donchin (2006); Salvaris and Sepulveda (2009)) or imagining a particular form of body movement (Hill et al., 2006). The system then infers which particular mental state a user was in: did you imagine moving your left hand or your right hand? Were you attending to the letter A or the letter S? The detection of these mental states is typically done using an appropriately trained classifier; for instance, a classifier trained on brain data representing three distinct classes of mental states: imagined left-hand movement, imagined right-hand movement or no-movement. This classifier is applied to single-trial measurements of a user's brain activity, typically measured using EEG, functional near-infrared spectroscopy (fNIRS) (Coyle et al., 2007), magnetoencephalography (van Gerven et al., 2009), fMRI (Ramsey et al., 2006; Andersson et al., 2011) or invasive measurements such as ECoG (Valderrama et al., 2012).

As interest in BCIs has increased, so has the breadth of potential applications of the BCI loop. Perceptual learning is a good example of a domain that has the potential for successful BCI implementations, given the well-established forms of brain activity associated with perceptual discrimination abilities. A BCI paradigm based on real-time, single-trial measurements of the MMN response could serve to augment existing educational settings related to second-language learning and musical training. The development of such a BCI represents the primary goal of this thesis. Similarly to the detection of P300 responses in BCI-speller paradigms, an appropriate classifier could be trained to detect whether an MMN response has occurred in single-trial measurements of AEPs, thus indicating whether a user has perceived a given sound contrast or not. In turn, this information could be used by instructors to determine whether specific materials are challenging for individual students, to adapt the difficulty of specific exercises and to provide students with direct feedback on their brain responses to the trained sounds.

## Neurofeedback

While the application of machine learning techniques to neuroimaging data is relatively recent, there has been a long-standing interest in monitoring brain activity in real-time. Since the 1970s the field of neurofeedback has been developing techniques that allow users to monitor measurements of their own brain activity in the form of an auditory or

visual signal. This enables the users to explore different strategies for modulating the targeted brain activity, usually with the goal of inducing various states such as relaxation or concentration, or regulating specific types of abnormal brain activity (Hammond, 2011).

Neurofeedback approaches are similar to other types of feedback approaches in that they are based on the principles of reinforcement learning and operant conditioning (Hammond, 2011; Shanks, 2010; Miltenberger, 2011). Positive feedback on task performance serves to reinforce particular strategies, while negative feedback prompts a modification in strategy. In the case of neurofeedback, participants receive near-instantaneous feedback on specific types of brain activity associated with a target mental state, allowing users to learn the relationship between self-induced changes in mental state and the feedback signal.

The majority of neurofeedback approaches have made use of quantitative EEG measurements (QEEG) of the power in different frequency bands from various scalp locations: delta ($1-4Hz$), theta ($4-8Hz$), alpha ($8-12Hz$), beta ($12-20Hz$), mu ($18-22Hz$) and gamma ($>25Hz$) (Hammond, 2011). The power of these measurements corresponds to the amount of synchronized oscillatory activity at those frequencies in different brain structures and networks (Bastiaansen et al., 2011). Increases in alpha power, for instance, are associated with the 'idling' activity of the visual cortex when a person closes their eyes, and are powerful enough that they can be visually identified in raw EEG traces, while activity in the theta band has been heavily linked to working memory processes (Jensen and Tesche, 2002; Lisman and Jensen, 2013). While an extensive description of oscillatory forms of brain activity is beyond the scope of this introduction, it suffices to say that it is likely that they play a fundamental role in structuring the activities of the central nervous system (Buzsáki, 2006; Luck and Kappenman, 2011).

EEG-based neurofeedback approaches are employed in wide variety of clinical settings, including the treatment of ADHA, depression, anxiety, brain injury and sleep disorders (Hammond, 2011). It has been estimated that over 100,000 people underwent neurofeedback-based treatments between 2000 and 2010 in the United States (New York Times, Oct 5 2010). However, the field has failed to produce conclusive empirical evidence that such treatments are in fact more effective than placebo treatments (Lansbergen et al., 2011). As such, a great deal of skepticism in such approaches remains: for instance, a conglomerate of Dutch health insurers decided in 2008 to stop reimbursement of costs incurred during neurofeedback-based treatments.

While questions remain regarding the effectiveness of traditional EEG-based neurofeedback, other approaches related to the treatment of specific physiological disorders such as tinnitus (Weisz et al., 2011) and those based on real-time fMRI measurements of activity in targeted brain regions (Yoo et al., 2006; Johnston et al., 2010; Zotev et al., 2011) have more clearly demonstrated the benefits of such methods. This may be due to the fact that the forms of brain activity targeted in these paradigms are more reliably linked to specific conditions and mental states than the oscillatory activities targeted by traditional paradigms.

The use of decoding methods in a neurofeedback paradigm allows targeting specific forms of brain activity, such as the MMN response, that are not typically measured in real-time. In turn, such methods can be employed in neurofeedback applications for healthy users intended to improve perceptual skills. A recent study investigating such a paradigm showed improvements in participants' ability to visually discriminate stimuli with different orientations following a 5-10 days of neurofeedback training based on the decoding of brain

activity using fMRI measurements from visual cortex regions associated with the perception of similar orientation stimuli (Shibata et al., 2011). A similar EEG-based approach centered on the decoding of the MMN response could offer the possibility of enhancing auditory perception of challenging sound contrasts such as foreign language phonemes or difficult-to-perceive musical sounds, such as out-of-tune notes. The development of such an approach is the primary focus of the work presented in the subsequent chapters.

## The Present Work

The essential question addressed by this dissertation can be summarized as follows: is it possible to modulate the brain responses underlying auditory perceptual discrimination using decoded-EEG neurofeedback in a manner that leads to marked improvement in perceptual skills? This question is investigated in a series of experiments and data analyses. This research is novel in the sense that it extends previous EEG-based neurofeedback approaches into the domain of auditory perceptual learning through the use of pattern classification methods.

Chapter Two presents the results of an experiment that investigated differences in the perception of English language phonemes by native speakers and highly-proficient second language (L2) speakers (Dutch natives). These two groups' categorical perception of a stop-consonant continuum (/pa/-/ba/) was assessed using both ERP and behavioral measures. It was found that the two groups differed not only with respect to their behaviorally measured category boundary, but also with respect to their brains' responses to sounds both within the category and at the category boundary. The results of this study provide a reference point with respect to both the individual behavioral and brain-response features that differ in native and L2 speech perception.

The results of an initial multivariate analysis of the EEG data from Chapter Two are presented in Chapter Three. Here, a series of within- and cross-participant classification analyses are performed using a binary regularized linear logistic regression algorithm. The results of these analyses are interpreted with respect to both individual differences in behavioral and brain responses and to group level differences. A complimentary set of analyses clarifies the relationship between the spatiotemporal features present in the dataset, the corresponding classifier weighting matrices, and specific components of the ERPs. It also provides a benchmark for the performance of classification methods that serve as the basis of the neurofeedback paradigm explored in the subsequent chapter.

The effectiveness of the proposed neurofeedback method in modulating specific components of the auditory evoked potential and enhancing corresponding perceptual discrimination abilities is investigated in Chapter Four. Participants completed four experimental sessions, all of which included a neurofeedback portion and listening tests investigating changes in their ability to discriminate the frequencies of simple auditory stimuli. These stimuli were preferred to speech stimuli due to the well-established behavioral and electrophysiological parameters associated with perceptual learning in this domain (Micheyl et al., 2006; Menning et al., 2000). A matched set of control participants underwent the same procedure but were provided with a form of sham feedback. Although no differences in the groups' discrimination sensitivities were observed, a corresponding difference in the brain responses measured during the feedback sessions was found. Such a result provides a 'proof-of-concept' for the neurofeedback paradigm, and suggests subsequent modification

of the approach that could lead to the enhancement of auditory perceptual learning.

In Chapter Five, a methodological framework for real-time tracking of perceptual discrimination on the basis of single-trial decoding of auditory evoked responses is presented. The method is illustrated using two datasets collected during the development of the neurofeedback paradigm, and is based on the perception of simple auditory stimuli in an MMN paradigm. Specifically, it shows how the output of a logistic regression classifier can be interpreted from a probabilistic perspective, such that it can be represented as ongoing changes in continuous value. This value in turn can be used to control a feedback signal, or alternatively, to monitor changes in specific cognitive states 'online'. The chapter also presents methods for optimizing various aspects of the decoding analysis, including preprocessing steps and data collection methods, as well as various approaches to the evaluation of features and generalization performance.

Future directions for this line of research are discussed in Chapter Six, along with the most challenging issues it faces. Specifically, potential changes are proposed for the stimulus materials, development and application of the classification methods, task design and participant/user base. The goal of such changes is to enrich the perceptual learning context of the neurofeedback procedure. Difficult issues such as poor-single trial performance, 'BCI-illiteracy' and response habituation are also discussed in the context of future research goals.

# 2

# Effects of native language on perceptual sensitivity to phonetic cues

*based on Brandmeyer, A., Desain, P. & McQueen, J.M., 2012. Effects of native language on perceptual sensitivity to phonetic cues. NeuroReport, 23(11), 653-657.*

**Abstract:** The present study used electrophysiological and behavioral measures to investigate the perception of an English stop consonant contrast by native English listeners and by native Dutch listeners who were highly proficient in English. A /ba/-/pa/ continuum was created from a naturally-produced /pa/ token by removing successive periods of aspiration, thus reducing the voice onset time (VOT). Whereas aspiration is a relevant cue for distinguishing voiced and unvoiced labial stop consonants (/b/ and /p/) in English, prevoicing is the primary cue used to distinguish between these categories in Dutch. In the electrophysiological experiment, participants listened to oddball sequences containing the standard /pa/ stimulus and one of three deviant stimuli, while the mismatch negativity (MMN) response was measured. Participants then completed an identification task on the same stimuli. Results revealed that native English participants were more sensitive to reductions in aspiration than native Dutch participants, as indicated by shifts in the category boundary and by differing within-group patterns of MMN responses and larger mean evoked potential amplitudes in the native English group for two of the three deviant stimuli. This between-group difference in the sensorineural processing of aspiration cues indicates that native language experience alters the way in which acoustic features of speech are processed in the auditory brain, even following extensive second-language training.

# Introduction

Speech perception is shaped by language experience. We can easily identify and discriminate different phonemes in our native language, but find this more difficult when it involves foreign language phonemes. The classical example of this phenomenon is the difficulty faced by speakers of Japanese in discriminating the English /r/ and /l/ sounds (Goto, 1971), but it has now been demonstrated for a variety of language pairs (Keating et al., 1981; Werker and Lalonde, 1988; Cutler et al., 2004; Strange, 1995). These changes in perceptual sensitivity appear to be learned. Developmental research has shown that, up until around 6 months of age, children can discriminate phonemes from native and non-native languages equally well (Best, 1993). However, between 6 and 12 months of age, children start to display a native-language bias (Kuhl et al., 1992, 2006). The cues learned during this formative stage of development persist into adulthood, systematically affecting the way in which foreign speech is perceived (Iverson et al., 2003).

Differences between native and non-native speakers are also reflected in their sensorineural responses during electroencephalographic (EEG) measurements of participants' brain activity while listening to speech sounds. The mismatch-negativity (MMN) component of the auditory evoked potential provides an index of auditory sensory discrimination ability (Näätänen et al., 2005; May and Tiitinen, 2009), and has been used to identify the neurophysiological correlates of speech perception (Näätänen, 2001). The MMN is typically observed in an unattended oddball paradigm as an increased fronto-central negativity peaking 100-300 ms following the presentation of a low-probability deviant stimulus in a sequence of high-probability standard stimuli. Stimulus pairs representing distinct phonemic categories in one's native language elicit a clear MMN response (Näätänen et al., 1997). Conversely, pairs representing phonemic categories not distinguished by one's native language have been found to elicit either a reduced response (Winkler et al., 1999) or none at all (Sharma and Dorman, 2000).

The present study examines electrophysiological and behavioral responses to English voiced and voiceless stop consonants by English and Dutch listeners. Whereas the presence of aspiration serves to distinguish /p/ from /b/, /t/ from /d/, and /k/ from /g/ in English, voiced and voiceless consonants are always unaspirated in Dutch. Instead, Dutch stops are primarily distinguished by the presence of prevoicing in only the voiced tokens (Mees and Collins, 1982; van Alphen and Smits, 2004). Speakers of these two languages must thus have learned to discriminate these sounds on the basis of distinct cues. However, the large majority of native-Dutch speakers are also highly proficient speakers of English, having completed up to six years of English language courses as part of their secondary education. This extensive second language (L2) training may have led to the same level of perceptual sensitivity to the English language cues as that shown by native speakers. This study investigates whether non-native speakers with extensive L2 training show the same sensorineural (MMN) responses as native speakers.

A single, natural speech token of the consonant-vowel (CV) syllable "pa" recorded by a native speaker of English was manipulated to form a /pa/-/ba/ continuum with successively reduced periods of aspiration, i.e. reduced voice onset time (VOT). These stimuli were used in an oddball paradigm during electrophysiological measurements, with the original /pa/ serving as a standard and three of the reduced aspiration stimuli serving as deviants. Behavioral responses to the stimuli were also evaluated using an identification task. If initial language learning dominates L2 learning, then it would follow that the

native-Dutch speakers will show reduced sensitivity to the acoustic cues present in the English language stimuli. However, if long-term L2 training is sufficient for perceptual learning of overlapping non-native acoustic cues, then the two groups should not differ in their sensorineural responses.

# Methods

## Participants

Thirteen native Dutch speakers (age range 19-35, 7 female), and thirteen native English speakers (age range 19-29, 7 female) reporting normal hearing completed the experiment. Two participants from each group were excluded due to excessive movement artifacts. The experiment was conducted primarily in the native language of the participants. All of the native Dutch speakers spoke English as L2, having completed at least 6 years of courses. The experiment was conducted with the approval of the local ethical committee, with all participants providing informed consent.

## Stimuli

A male, native English speaker (the author) recorded repetitions of the syllable /pa/ in a sound-proof room at 44.1 kHz, 16-bit quality using a MOTU 828mk2 audio interface and a Bruell & Kjaer 4006 microphone. One token with a duration of approximately 450 ms was selected.

Six additional stimuli were created by removing successive 11 ms sections of the aspirated portion of the original /pa/ (i.e. reducing VOT). The duration of the syllable as a whole was preserved by inserting additional periods of voicing during the steady-state portion of the vowel. This led to a total of seven stimuli: the original /pa/ token, with an 85 ms VOT (the standard), and six manipulated stimuli with varying VOTs: 74 ms, 63 ms, 52 ms, 41 ms, 30 ms, and 19 ms. This continuum of stimuli sounded progressively more like /ba/ to native-English speakers, as seen post hoc in the behavioral results.

## Electrophysiological Measurements

Three of the six manipulated stimuli [19 ms, 41 ms, and 63 ms VOTs] were used as deviants for presentation in oddball sequences during EEG recording. In each of these conditions, one deviant was presented with a 15% likelihood amidst presentations of the standard stimulus (85 ms VOT), with an onset-to-onset inter-stimulus interval of 1200 ms. These sequences were presented in blocks lasting approximately 6.5 minutes, each containing 45 repetitions of one of the deviants. All participants completed at least two blocks for each of the three deviants; three English and two Dutch participants completed three blocks for each deviant.

EEG was measured at a sample rate of 512 or 2048 Hz inside a shielded cabin using a BioSemi ActiveTwo amplifier with 64 Ag/AgCl electrodes and the international 10-20 system for electrode placement. During measurement, participants viewed a self-selected silent film on a computer screen while stimuli were presented over Monocor MKS-28 stereo loudspeakers at approximately 70 dB SPL. There was a short break between blocks during

which the experimenter asked the participants short questions about the films in order to ensure that participants were alert and attending to the films.

The EEG data was then resampled to 128 Hz, referenced to the averaged mastoid leads, and band-pass filtered between 1-30 Hz. Data from electrodes with offsets that exceeded $\pm 35\ mV$ during recording were replaced using the average of the neighboring electrodes. Epochs with artifacts exceeding $\pm 75\ \mu$V at any electrode location were discarded. The mean offset between -50 ms and 0 ms relative to stimulus onset was used as a baseline value for each epoch. Individual grand-average ERPs were calculated for the standard and three deviant stimuli using data from nine fronto-central electrode locations: F1, Fz, F2, Fp1, FpZ, Fp2, C1, Cz, and C2. MMN difference waves were produced by subtracting the ERP for the standard stimulus from that of each of the deviants. Based on the observed peaks in the group-level difference waves, the peak latency of the MMN between 250 and 350 ms was determined for each participant in each condition, and the mean MMN amplitude was calculated in a 50 ms window around the peak.

### Behavioral Measurements

Participants were seated in a sound-proof cabin and presented with one of the seven stimuli at approximately 80 dB SPL using Sony MDR-7506 headphones on each trial. They were asked to judge whether it was /pa/ or /ba/. Responses were registered using a Multimedia Systems touchscreen. Each stimulus was presented a total of 20 times in random order.

### Statistical Analysis

For both native-language groups, one-sample t-tests were used to establish whether mean MMN amplitude in each of the three deviant conditions significantly differed from zero. Electrophysiological and behavioral data were analyzed using repeated measures ANOVAs, with VOT as a within-subjects factor, and native language as a between subjects factor. Within-group comparisons of VOT conditions were analyzed using paired-samples t-tests, while between-group differences were analyzed per condition using one-way ANOVAs. When applicable, results were corrected using Greenhouse-Geisser estimates of sphericity.

## Results

### Identification Task

Identification performance is summarized in Fig. 2.1. An effect of VOT ($F(6, 15) = 116.315, p < .001$) and an interaction of VOT and native-language ($F(1, 20) = 4.294, p = .02$) were found. The effect of native-language was marginally significant ($F(1, 20) = 3.969, p = .06$). Native-English speakers were more likely to categorize stimuli as /ba/ in the 30 ms VOT ($F(1, 20) = 5.023, p < .05$) and 41 ms VOT ($F(1, 20) = 5.080, p < .06$) conditions than the native-Dutch speakers. For the 19 ms VOT stimulus, the difference was only marginally significant ($F(1, 20) = 4.230, p < .06$).

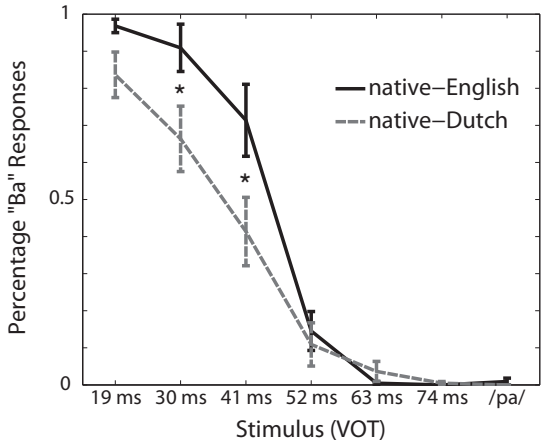


Figure 2.1: *Categorization results for native-English and native-Dutch participants. Stimuli for which a significant between-groups difference was found ($p < .05$) are indicated by an asterisk.*

## Electrophysiological Tests

Grand Average ERP plots for both groups at fronto-central electrode locations can be seen in Fig. 2.2. As VOT was decreased in the three deviant stimuli, a corresponding increase in the negativity of the N2 component for both groups was observed. Previous studies investigating electrophysiological responses to speech stimuli using a VOT continuum have shown systematic variations in the latencies of negative peaks in the ERPs related to the VOT of a given stimulus (Sharma and Dorman, 1999, 2000). In order to assess whether the differences in the VOT between the standard and deviant stimuli would affect the calculation of the MMN difference wave, the peak latencies of the N1 and N2 component of the ERPs were analyzed. No significant effects of VOT or native-language were found for either component.

For the N1 component, mean peak latencies were 132 ms (SD = 12 ms) for the 85 ms VOT stimulus, 134 ms (SD = 12 ms) for the 63 ms VOT stimulus, 129 ms (SD = 8 ms) for the 41 ms VOT stimulus, and 132 ms (SD = 14 ms) for the 19 ms VOT stimulus. The mean peak latencies of the N2 component were 305 ms (SD = 13 ms) for the standard /pa/ stimulus (85 ms VOT), 309 ms (SD = 14 ms) for the 63 ms VOT deviant, 304 ms (SD = 15 ms) for the 41 ms VOT deviant, and 303 ms (SD = 15 ms) for the 19 ms VOT deviant.

The main analysis focused on between-groups effects. Mean MMN amplitude was significantly affected by both VOT ($F(2, 19) = 27.592, p < .001$) and native-language ($F(1, 20) = 8.459, p < .01$). Native-English speakers showed larger mean MMN amplitudes in the 41 ms VOT ($F(1, 20) = 12.351, p < .001$) and the 63 ms VOT deviant conditions ($F(1, 20) = 5.826, p < .05$), but not in the 19 ms VOT condition. Thus, the MMN responses of the two groups differed only for deviants near the category boundaries found in the behavioral results. The analysis of MMN peak latency data revealed no significant effects of VOT or native language.

Additional analyses verified whether an MMN response significantly differing from zero was observed for each of the language groups to each deviant. MMN difference waves

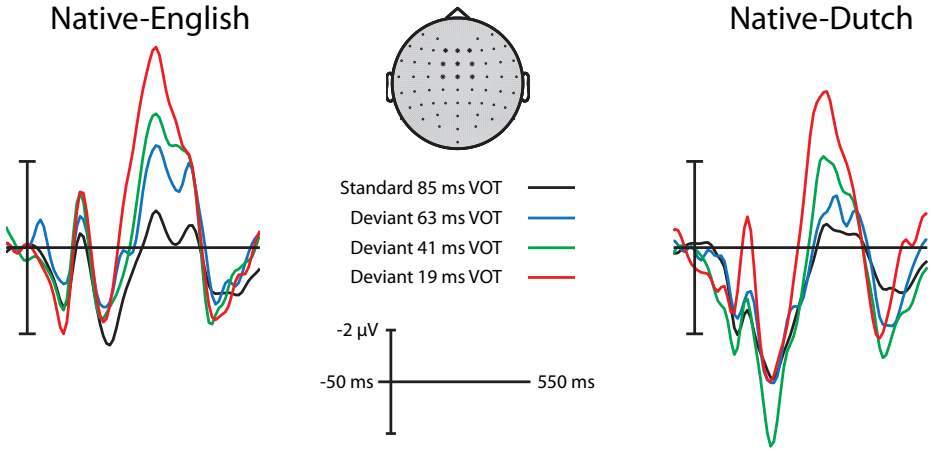


Figure 2.2: *Grand Average ERPs for native-English and native-Dutch listeners at fronto-central electrode locations. Responses for all four stimuli conditions are overlaid at each electrode location. Electrode locations are indicated by large dots on the scalp map, with the same relative positioning as in the ERP plots.*

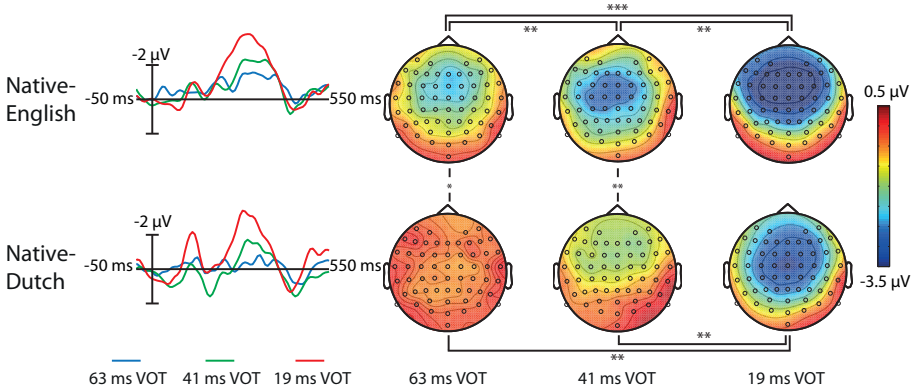


Figure 2.3: *Grand-average MMN difference waves at fronto-central electrode locations, along with average MMN scalp distributions for the 50 ms window surrounding the group peak for the three deviant stimulus conditions. Statistical effects of mean MMN amplitude are illustrated using the scalp distributions. For each of the two groups, brackets connecting two conditions indicate significant within-subjects effects of conditions, as revealed by a paired-samples t-test. Between-groups effects, based on one-way ANOVA results, are illustrated by lines connecting the two participant groups in a given condition. Statistical significance levels are indicated by asterisks:* $* = p < .05, ** = p < .01, *** = p < .001$

and scalp topographies for both groups across conditions are plotted in Fig. 2.3. For the native-English speakers, a significant MMN response was found for all three deviants. Mean MMN amplitude was -4.32 $\mu$V for the 19 ms VOT deviant ($t_{10} = -12.061, p < .001$), -2.99 $\mu$V for the 41 ms VOT deviant ($t_{10} = -12.963, p < .001$), and -1.91 $\mu$V for the 63 ms VOT deviant ($t_{10} = -9.429, p < .001$). Mean MMN amplitude in the 19ms condition was larger than both the 41 ms VOT ($t_{10} = 3.338, p < .01$) and the 63 ms VOT conditions ($t_{10} = 7.763, p < .001$). Mean amplitude in the 41 ms VOT condition was also larger than in the 63 ms VOT condition ($t_{10} = 3.867, p < .01$). The peak latency of the MMN difference wave was 303 ms (SD = 30 ms) for the 19 ms VOT deviant, 309 ms (SD = 33 ms) for the 41 ms VOT deviant, and 323 ms (SD = 25 ms) for the 63 ms VOT deviant. No significant between condition effects were found for the peak latency measurements.

Results for the native-Dutch speakers also revealed MMN responses significantly differing from zero for all three deviants: -3.63 $\mu$V for the 19ms VOT deviant ($t_{10} = -6.568, p < .001$), -1.75 $\mu$V for the 41 ms VOT deviant ($t_{10} = -6.622, p < .001$), and -0.93 $\mu$V for the 63 ms VOT deviant ($t_{10} = -2.631, p < .05$). Across conditions, mean MMN amplitude was larger in the 19 ms VOT condition than in the 41 ms VOT ($t_{10} = 4.284 p < .01$) and 63 ms VOT conditions ($t_{10} = 4.451, p = .001$). The peak latency of the MMN difference wave was 300 ms (SD = 34 ms) for the 19 ms VOT deviant, 310 ms (SD = 24 ms) for the 41 ms VOT deviant, and 310 ms (SD = 34 ms) for the 63 ms VOT deviant. The differences in peak latencies between the three conditions were not significant.

## Discussion

There was a clear difference in perception between the English- and Dutch-speaking listeners. In the categorization task, a clear shift in the category boundary for non-native speakers is evident, with the boundary occurring one step earlier (between 41 and 52 ms VOT) for English than for Dutch speakers (between 30 and 41ms VOT). There was also a sharper rise in the categorization function for native speakers. These findings are in line with previous results on the perception of stops (Liberman et al., 1961; Pisoni, 1973). As would be expected on the basis of Dutch and English phonology, Dutch listeners are less sensitive to the presence of aspiration than English listeners.

These differences in sensitivity to aspiration were also reflected in the electrophysiological data. While MMN responses were observed for both groups in all three deviant conditions, native-English participants exhibited significantly larger responses than native-Dutch participants in the 41 and 63 VOT conditions. In the case of the 41ms VOT deviant, the MMN response of native-English speakers was significantly larger than their response to the 63 ms deviant, corresponding to the category boundary observed in the behavioral measurements. This effect was shifted for the native-Dutch participants, with a significantly larger response observed for the 19 ms VOT deviant as compared to the 41ms and 63 ms VOT deviants, again corresponding to the category boundary observed in the behavioral tasks.

Previous research has shown that prevoicing serves as the most important cue in distinguishing voiced and voiceless stops in Dutch (van Alphen and Smits, 2004). However, 25% of the voiced stops in a production experiment did not contain any prevoicing, and when these stimuli were used in a perceptual experiment, they were much more likely to be judged as voiceless than stimuli containing prevoicing (van Alphen and Smits, 2004). Van

Alphen and Smits note that the Dutch language may thus be undergoing sound change as a result of the influence of English, through both secondary language education and the abundance of English language media in the Netherlands. Here, however, we find that our Dutch participants, in spite of their L2 fluency, are still more likely to perceive the ambiguous stimuli as voiceless than native-English participants. If sound change is taking place, it is not yet complete. Native-language speech categories continue to determine sensorineural responses to auditory cues, and hence to color the perception of even highly proficient L2 speakers.

# Conclusion

In sum, a decreased sensitivity, as indexed by the MMN response to aspiration cues in English stimuli, was observed for native Dutch speakers relative to native English speakers, and which corresponded to differences in the categorical functions observed for the two groups in an identification task using the same stimuli. These results suggest that experience in infancy with one's native language causes long-term changes in the nervous system that affect the acoustic analysis of speech, and are even more striking given the native Dutch participants' extensive experience with English.

# 3

# Decoding speech perception by native and non-native speakers using single-trial electrophysiological data

*based on Brandmeyer, A., Farquhar, J., McQueen, J.M. & Desain, P., 2013. Decoding speech perception by native and non-native speakers using single-trial electrophysiological data. PLOS One, 8(7), e68261.*

**Abstract:** Brain-computer interfaces (BCIs) are systems that use real-time analysis of neuroimaging data to determine the mental state of their user for purposes such as providing neurofeedback. Here, we investigate the feasibility of a BCI based on speech perception. Multivariate pattern classification methods were applied to single-trial EEG data collected during speech perception by native and non-native speakers. Two principal questions were asked: 1) Can differences in the perceived categories of pairs of phonemes be decoded at the single-trial level? 2) Can these same categorical differences be decoded across participants, within or between native-language groups? Results indicated that classification performance progressively increased with respect to the categorical status (within, boundary or across) of the stimulus contrast, and was also influenced by the native language of individual participants. Classifier performance showed strong relationships with traditional event-related potential measures and behavioral responses. The results of the cross-participant analysis indicated an overall increase in average classifier performance when trained on data from all participants (native and non-native). A second cross-participant classifier trained only on data from native speakers led to an overall improvement in performance for native speakers, but a reduction in performance for non-native speakers. We also found that the native language of a given participant could be decoded on the basis of EEG data with accuracy above 80%. These results indicate that electrophysiological responses underlying speech perception can be decoded at the single-trial level, and that decoding performance systematically reflects graded changes in the responses related to the phonological status of the stimuli. This approach could be used in extensions of the BCI paradigm to support perceptual learning during second language acquisition.

# Introduction

Learning foreign languages is difficult, in part because they often make use of sounds which are unfamiliar. Moreover, foreign speech sounds can be difficult to discriminate from one another, depending on the types of phonemes used in one's native language. Studies of human language perception have made use of EEG measurements to reveal differences in the processing of speech sounds by the brains of native and non-native listeners. The results of these studies are typically based on the analysis of event-related potentials collected over hundreds of trials and using many individual participants. This is done because the signals of interest are much smaller in amplitude than the ongoing brain activity measured during single-trials (Handy, 2005). By averaging EEG data collected during repeated time-locked presentations of speech sounds, brain activity unrelated to the stimulus presentation is eventually cancelled out, leaving only the brain's responses to the speech sound. But what if it were possible to detect these signals in single-trial EEG data? Research using multivariate pattern classification methods and brain-computer interface (BCI) paradigms has shown that this is feasible for signals such as the P3 response (Farwell and Donchin, 1988; van Gerven et al., 2009). In turn, users are able to control different types of systems (e.g. communication devices and computers) using mental activity alone by, for instance, attending to items in a flashing menu. If it was also possible to detect the brain responses underlying speech perception, it could allow for the development of BCIs that support the learning of foreign languages through the monitoring of ongoing perception, or by providing feedback to users on their brain's responses.

To this end, a study was conducted using a multivariate analysis of EEG data collected during passive auditory perception of English language phonemes by native and non-native speakers of English. It investigated whether such methods are sensitive to the different electrophysiological response patterns elicited when native and non-native listeners are presented with pairs of stimuli from a continuum of phonemes representing either within- or across-category contrasts. Additionally, the study used the same methods in conjunction with two cross-participant data sets to address questions regarding the consistency of the functional brain organization underlying speech perception across individuals both within and between language groups.

## The mismatch negativity component and research on speech perception

Previous research using auditory event-related potentials (ERPs) has revealed consistent differences between native and non-native speakers in the brain responses underlying the perception of phonetic contrasts (Näätänen et al., 1997; Winkler et al., 1999; Sharma and Dorman, 2000). These findings are often based on analysis of the mismatch negativity (MMN) component of the auditory ERP, which is typically seen at fronto-central scalp locations following the presentation of a low-probability 'deviant' stimulus. As the MMN is typically elicited using a passive listening paradigm, it is thought to provide a pre-attentive index of perceptual discrimination abilities (Näätänen et al., 2007; Duncan et al., 2009). In addition, the amplitude and latency of the MMN have been shown to be modulated by the stimulus contrast employed. Large differences between standard and deviant stimuli will lead to increases in MMN amplitude as well as a decrease in its latency, while smaller differences will reduce the amplitude and increase the latency (Kujala et al., 2007).

The MMN has been observed in response to both changes in acoustic features of phonemes typical of within-category variation (Sharma and Dorman, 1999; Rivera-Gaxiola et al., 2000; Kasai et al., 2001) as well as when presenting stimulus contrasts representing two distinct phonemic categories (Näätänen et al., 1997; Dehaene-Lambertz, 1997; Winkler et al., 1999; Sharma and Dorman, 1999; Phillips et al., 2000; Sharma and Dorman, 2000). A comparison of the MMN responses evoked by stimuli from a phonetic continuum containing both within- and across-category deviants showed that across-category responses were significantly larger than within-category responses (Sharma and Dorman, 1999). When non-native listeners are presented with a meaningful phonetic contrast in an unfamiliar language, the measured ERPs typically show a reduced (Näätänen et al., 1997) or absent MMN response (Dehaene-Lambertz, 1997; Winkler et al., 1999; Sharma and Dorman, 2000) relative to native speakers. Thus it would seem that MMN responses observed in response to phonemes show a graded effect, with respect to both the categorical status of the phonetic contrast as well as to the linguistic background of individual listeners.

While MMN responses to artificial tone stimuli are consistently reported in the N1 interval (Näätänen et al., 2007), studies using phonetic stimuli have reported MMN in both the N1 (Näätänen et al., 1997; Kasai et al., 2001) and N2 (Sharma and Dorman, 1999, 2000) intervals. It has been suggested that stimulus contrasts representing distinct phonetic categories give rise to changes in the N2-P3 complex of the auditory ERP, while effects in the N1 interval reflect the processing of acoustic differences in the stimuli (Maiste et al., 1995). Other findings have also suggested a distinction between early and late MMN responses to speech (Korpilahti et al., 1995) and speech-like (Korpilahti et al., 2001) stimuli. Additionally, the same auditory oddball paradigms used to elicit MMN responses have also been shown to modulate mid-latency components prior to the N1 (Grimm et al., 2011), and to elicit a negative component following the P3a response known as the reorienting negativity (RON) (Schroger and Wolff, 1998; Munka and Berti, 2006). As such, depending on stimulus and sequence parameters, ERPs collected on deviant trials during MMN measurement paradigms can be expected to show an enhancement of negative components in one or more time intervals relative to standard trials. The question addressed by the present study was whether these (or other) components could be detected reliably at the single-trial level.

## Multivariate analysis methods and auditory perception

While the neurophysiology of speech perception has been examined extensively using traditional ERP methodologies, there has recently been an increasing interest in the use of multivariate pattern classification methods to address questions regarding the functional organization of cognitive processes using data collected at the single-trial level (Haynes and Rees, 2005a; Kamitani and Tong, 2005; Haynes and Rees, 2006; Formisano et al., 2008; Kay et al., 2008), and to develop BCIs based on real-time measurements of brain activity. Whereas the traditional univariate methods used to analyze neurophysiological signals such as the BOLD response or ERP measurements focus on amplitude differences at individual data points (i.e. sensors, time points, voxels), multivariate methods are sensitive to differences in the distribution of responses across high-dimensional feature spaces. Moreover, when used with data collected at the single-trial level, additional information contained within the single-trial responses is available which might otherwise be lost when

averaging across trials.

Several BCI studies have used multivariate methods to detect different classes of auditory ERPs elicited by target and non-target stimuli in an active task. Such tasks are know to elicit a P3 response (Polich, 2007; Duncan et al., 2009), and have also been used with stimuli in the visual (Farwell and Donchin, 1988; van Gerven et al., 2009) and tactile (van der Waal et al., 2012) modalities. Halder and colleagues reported on a system capable of making binary choices using auditory targets which differed in either loudness, pitch or direction (Halder et al., 2010). Systems capable of distinguishing a larger number of classes using either spatial (Schreuder et al., 2010; Belitski et al., 2011) or a combination of spatial and frequency (Höhne, 2011) cues have also been reported. Additional work has shown that the use of speech stimuli can enhance classifier performance relative to artificial stimuli (Höhne et al., 2012). While the principal focus in these studies has been the elicitation of a P3 response for use as a control signal in determining whether a target stimulus has been presented, some of the studies just mentioned have also reported on the contribution of negative ERP components in the 100-300 ms post-stimulus onset time interval to overall BCI performance (Schreuder et al., 2010; Höhne, 2011).

Multivariate approaches have also been used in several studies to investigate auditory perception of speech and music at the single-trial level. In the music domain, it has been shown that decoding perceived music from EEG data at the single trial level is possible, and that decoding using cross-participant data sets leads to similar overall performance as compared to within-participant analyses (Schaefer et al., 2011). Additional work using EEG data has shown that the decoding of accented vs. unaccented beats in an isochronous sequence is possible, during both active perception as well as during a subjective-accenting task, and that decoding performance generalizes across these conditions (Vlek et al., 2011). With regard to speech perception, it has been shown that the brain activity underlying the perception of different vowels and different speaking voices can be decoded from single-trial fMRI data (Formisano et al., 2008). A recent study by Herrmann and colleagues demonstrated that both unexpected changes in low-level acoustic features as well as syntactic-rule violations can be also decoded using MEG data, with cross-participant analyses showing a high-degree of consistency in both the spatial distribution of features as well as in overall performance relative to individual analyses (Herrmann et al., 2012).

## The present study

This study aims to extend these findings by examining whether the perception of phonetic contrasts representing within- or across-category contrasts can be decoded using single-trial EEG data. This is accomplished using the dataset from Chapter Two, which addressed within- and between-group differences in the perception of a phonetic continuum by native (English) and non-native (native-Dutch) speakers (published as Brandmeyer et al., 2012). This makes it possible to interpret the results of the present classification analyses with respect to outcomes of traditional ERP analyses as well as individual behavioral measurements. In addition to the within-participant analyses, the results of both multi-trial and cross-participant decoding analyses are presented, and, on the basis of these results, the potential for novel extensions of the BCI paradigm to the domain of second language learning is discussed.

| | |
|---|---|
| Participants | 11 Native-English speakers, 11 Native-Dutch speakers |
| Stimuli | English language CV (/pa/-/ba/) syllables: 85 ms VOT (standard), 63 ms, 41 ms and 19 ms VOT (deviants) |
| Stimulus Intensity | approx. 70 dB |
| Stimulus Duration | approx. 450 ms |
| ISI | 1200 ms |
| Deviant Likelihood | 15 % |
| Trial Counts | 90-135 per deviant condition, per participant |
| EEG System | 64 Channel BioSemi Active2 + left & right mastoids, Horizontal & Vertical EOG |
| ERP analysis electrodes | F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2 |
| Sampling Rate | 512 Hz or 2048 Hz |

Table 3.1: *Details of experimental paradigm*

# Materials and Methods

## Ethics Statement

All participants provided written informed consent prior to their participation in the experiment. The experiment was performed in accordance with the guidelines of and was approved by the ethics committee of the Faculty of Social Sciences, Radboud University Nijmegen.

## Participants and Stimuli

The present study was a reanalysis of the data collected in Chapter 2 during passive speech perception of English language phonemes by native and non-native speakers of English. The non-native speakers who participated in the experiment were all native speakers of Dutch, and were also proficient speakers of English, having undergone at least 6 years of English language education. The native speakers are referred to as 'native-English' and the non-native speakers as 'native-Dutch'. Data for the same eleven participants in each of the two language groups as in the original study were used. A summary of the experimental design can be found in Table 3.1.

Four consonant-vowel (CV) syllables representing an English language stop consonant continuum were used as stimuli during the EEG measurements. A recording of the CV syllable /pa/ spoken by a male native-English speaker with a Voice Onset Time (VOT) of 85 ms was used to create the other three stimuli by removing successive 22 ms portions of the aspirated portion of the original recording prior to voice onset. Thus, the VOTs of these stimuli were 63 ms, 41 ms and 19 ms. The duration of these stimuli were preserved by inserting additional periods of voicing in the voiced portion of the recording. Waveforms of the four stimuli are presented in Figure 3.1.

The purpose of this manipulation was to produce a continuum which sounded progressively more like /ba/ to native speakers of English. In the English language, voiced and voiceless stop consonants (e.g. /d/ vs. /t/, /b/ vs./p/) are primarily distinguished from one another on the basis of VOT, while in the Dutch language the voiced and voiceless stop consonants are primarily distinguished by the presence of pre-voicing (Collins and Mees,

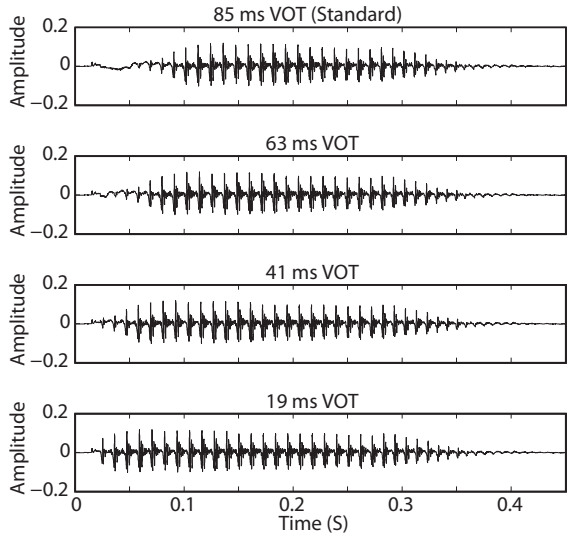


Figure 3.1: *Experimental stimuli waveforms. A recording of the English CV syllable /pa/ with a voice onset time of 85 ms was used as the standard stimulus during EEG recordings. The three deviant stimuli were created by removing successive 22 ms portions of the aspirated period prior to voice onset in the original 85 ms VOT standard stimulus, and by inserting additional periods of voicing to preserve the duration of each stimulus. The onset of the initial plosive burst was preserved for all of the stimuli.*

1996; van Alphen and Smits, 2004). Results of the original study indicated that both groups perceived the 63 ms VOT stimulus as /pa/ (within-category relative to the 85 ms VOT stimulus) and the 19 ms VOT stimulus as /ba/ (across-category). A between-groups difference was observed with respect to the 41 ms VOT stimulus, which was more likely to be perceived as /ba/ by native-English speakers and as /pa/ by native-Dutch speakers (Brandmeyer et al., 2012). In other words, the 41 ms VOT was located near each of the two groups' category boundaries, but fell on opposite sides.

During EEG measurements, these stimuli were presented in pseudorandom oddball sequences containing a standard stimulus (always the 85 ms VOT stimulus) and one of three deviant stimuli (see Table 3.1 for details of the oddball sequence parameters). There were three different EEG measurement conditions, which will be referred to subsequently using the name of the deviant stimulus which was used: '63 ms VOT deviant', '41 ms VOT deviant' and '19 ms VOT deviant'. During the EEG measurements, auditory stimuli were presented over loudspeakers while participants watched self-selected silent movies. Participants were instructed to ignore the auditory stimuli and to attend to the movie. This type of passive listening paradigm is typically used in conjunction with oddball stimulus sequences to elicit the MMN component of the auditory ERP (Näätänen et al., 2007; Duncan et al., 2009). After the passive oddball procedure, all participants completed a two-alternative forced-choice task (without EEG measurement) in which they actively identified the stimuli on the /ba-pa/ continuum.

## EEG Data Collection and Processing

Details of the EEG measurement system can be found in Table 3.1. Measurements were conducted inside a shielded electric cabin using a BioSemi ActiveTwo amplifier with 64 Ag/AgCl electrodes placed according to the international 10-20 system. Stimuli were presented to participants at approximately 70 dB SPL using a Monocor MKS-28 stereo loudspeaker system. Raw EEG data was measured along with left and right mastoid leads, horizontal and vertical EOG at a sample rate of either 512 or 2048 Hz. Filtering, referencing and additional preprocessing was performed offline, as described below.

For the present analysis, EEG data measured in each of the three deviant stimuli conditions were processed in non-overlapping epochs ranging from -200 ms before stimulus onset to 1000 ms post stimulus onset. Only epochs collected during trials containing a deviant stimulus and the standard trials immediately preceding them were selected for analysis, meaning an equal number of standard and deviant trials were analyzed in each condition. In each epoch, a spherical-spline interpolation procedure (Perrin et al., 1989) was used to repair individual EEG channels whose power in the 50Hz band exceeded 1000 $\mu V^2$ or whose offset exceeded $\pm$25 mV. An average of 2.86 channels were repaired per epoch (St. Dev. = 1.95). The data were then resampled to 128 Hz, and an independent component analysis (using the infomax ICA algorithm as implemented in the 'runica' function of the EEGLab toolkit (Delorme and Makeig, 2004)) was performed on each participant's data in order to identify and remove components containing non-EEG artifacts such as muscle or eye movements (Jung et al., 2000). Only components which accounted for more than 1 % of the overall variance in the data were considered for removal. For each of the components under consideration, the variance in each epoch of data was calculated. The mean variance across epochs was then calculated for each component. Components whose mean variance exceeded a threshold set to the average variance across all considered

components were then visually inspected to verify that their time course and topography were typical of non-EEG artifacts such as neck and eye movements (highly focal spatial distribution, large amplitude). Incremental adjustments to the threshold were then made on a per participant basis to ensure that components including non-artifactual activity were not removed. This approach is similar to that used in a previous analysis of individual auditory ERPs by Bishop and Hardiman (Bishop and Hardiman, 2010). An average of 5.14 components (St. Dev. = 2.01) were removed from each participant's data. Following the removal of these components, data were reprojected onto the measurement channels, and any epochs containing activity exceeding $\pm 75\ \mu V$ relative to the mean activity in the 100 ms window preceding stimulus onset were also removed from the dataset. On average, 97% of the analyzed epochs (St. Dev. = 3.7%) and at least 70 trials per stimulus in each of the three conditions remained following artifact rejection for all participants. Finally, data were band-pass filtered between 1 and 25 Hz, re-referenced to the average of the two mastoid leads, and baseline-corrected using the mean amplitude of the data in the 100 ms window preceding stimulus onset. All preprocessing was done using the Fieldtrip toolbox (Oostenveld et al., 2011) in MATLAB. All subsequent classification analyses made use of EEG data in the time range between 0 and 700 ms relative to stimulus onset.

## Classification Analyses

Data collected for both native-language groups in each of the three measurement conditions were analyzed using receiver operating characteristics (ROC) analysis. Typically used for problems in the domain of signal detection theory, ROC analyses are often used to analyze both the performance of classifiers (Fawcett, 2006) as well as the discriminability of feature distributions (Green and Swets, 1974; Schreuder et al., 2010). Here, we use area-under-the-ROC-curve (AUC) scores to quantify the separability of one-dimensional spatio-temporal feature distributions. These scores fall in the range of [0,1], with a score of .5 representing the no-discrimination line in the ROC graph.

Individual participant's single-trial EEG data (64 channels x 90 samples per epoch) were used to train a set of quadratically regularized linear logistic regression classifiers (Bishop, 2009). The regularization term is needed to limit the complexity of the classifier which prevents over-fitting in the high-dimensional input feature space (Farquhar and Hill, 2013). To find the optimal regularization strength (or equivalently classifier complexity), a simple grid search with strengths of [.001 .01 .1 1 10 100] times the total data variance was used, as empirically this range has been found to give high performance.

A series of within-participants analyses were carried out to determine whether differences in the perceived categories of pairs of phonemes influenced single-trial decoding performance. To this end, a separate analysis was performed using data collected in each of the three stimulus conditions: '63 ms VOT deviant', '41 ms VOT deviant' and '19 ms VOT deviant'. These names will be used subsequently to refer to each of the within participant analyses. All of the within-participant analyses investigated a binary comparison of single-trial EEG data collected during standard trials (always the 85 ms VOT stimulus) and deviant trials in a given measurement condition. A fourth analysis was performed which included all of the data collected across conditions for each individual participant. The results of this analysis were used to compare mean decoding performance for each of the four stimuli with the individual behavioral identification scores collected in Chapter Two using the same stimuli.

In each analysis, an equal number of epochs of data recorded during the presentation of a deviant stimulus and the standard stimulus immediately preceding it represented the two classes in a binary classification problem. On average, 202.5 consecutively recorded trials (St. Dev. = 42.2) were available for each of these classification analyses. All of the within-participant analyses utilized a ten-fold cross validation procedure, in which subsets of the available data were used for training and testing (90% and 10%, respectively) the classifier in each of the folds.

A subsequent analysis of the classifier decisions obtained at the single-trial level was performed in order to determine the performance benefits of using multiple trials. For this, classifier decisions obtained for all available data epochs in the test folds of the within-participant analyses conducted for the 19 ms VOT deviant condition were used. Each decision represents a continuous probability $p(+|x_i)$ that a given data epoch $x_i$ belongs to the target class $+$. In the context of a logistic regression classifier:

$$p(+|x_i) = 1/(1 + e^{-f(x_i|+)}) \tag{3.1}$$

Where, $f(x_i|+) = w^\top x_i + b$ is the classifier decision value given a set of classifier weights $w$ and a bias term $b$. For our analysis, these probabilities were combined for non-overlapping groups of $n$ consecutive data epochs $[x_1 \dots x_n]$ belonging to each of the two classes using a naive-Bayes formulation under the assumption of independence in the following manner:

$$P(+|x_1 \dots x_n) = \frac{\prod_1^n p(+|x_i)}{\prod_1^n p(+|x_i) + \prod_1^n (1 - p(+|x_i))} \tag{3.2}$$

Noting that for Logistic regression $1 - p(+|x_i) = e^{f(x_i|+)} p(+|x_i)$, the denominator becomes:

$$\prod_1^n p(+|x_i) + \prod_1^n e^{f(x_i|+)} p(+|x_i) = (\prod_1^n p(+|x_i))(1 + e^{\sum_1^n f(x_i|+)}) \tag{3.3}$$

and (3.2) becomes:

$$p(+|x_1 \dots x_n) = \frac{\prod_1^n p(+|x_i)}{\prod_1^n p(+|x_i)(1 + e^{\sum_1^n f(x_i|+)})} = \frac{1}{1 + e^{\sum_1^n f(x_i|+)}} \tag{3.4}$$

Thus, one can combine decisions by simply adding together classifier decision values, which is not only simpler but also less prone to numeric round-off errors.

Another aim of the present study was to investigate whether the decoding of categorical speech perception is possible across different individuals, both within and between native-language groups. Two additional classifiers were trained on cross-participant datasets collected in the 19 ms VOT deviant condition (70 consecutive trials per class for each participant). This stimulus contrast was chosen because it represented a clear categorical distinction for both native-English and native-Dutch listeners. The first classifier was trained using data from 10 of the 11 native speakers whose within-participant classification results were significantly above chance level (see below for details), and will be referred to with the name 'Cross-PP Native'. The second made use of all 22 participants' data, and will be referred to with the name 'Cross-PP All'.

Both cross-participant classifiers were trained using a double-nested cross-validation procedure in order to account for the additional inter-subject variability introduced by these

datasets. Such a procedure provides a means for selecting an optimal hyperparameter for a given classification problem whilst estimating generalization performance. In each main fold of the data, one participant's data served as a test set (for estimating cross-participant performance generalization), while the remaining participants' data formed the classifier training set. An additional set of nested folds repeated this procedure in order to estimate the regularization parameter, with the participant whose data was used for the test set being excluded from the nested analyses.

A final series of classification analyses were conducted that aimed to decode the native language (English or Dutch) of a given participant using either EEG or behavioral data. In the previous analyses, the labels assigned to the data used for training and testing the classifiers indicated whether an individual epoch was collected on a standard or deviant trial. Here, the labels indicated whether the data belonged to a native-English or native-Dutch speaker. The classifier performance levels obtained in such an analysis indicate the extent to which the response patterns (either EEG or behavioral) obtained from the two native-language groups generalize within-group, and how well these response patterns can be distinguished from one another at the group level.

Four separate analyses were performed with each of the following data sets: concatenated single-trial data from all three measurement conditions (70 total data segments per participant), concatenated grand average data from all three measurement conditions, concatenated grand average data measured from the 63 ms and 41 ms VOT deviant stimuli (the two stimuli for which a significant between groups difference in ERP responses was observed in Chapter Two), and the vector of behavioral identification scores for all 7 stimuli measured in the categorization task in Chapter Two. An additional analysis combined the single-trial predictions across trials on a per-participant basis in the same manner as previously described in equation 3.1. A naming scheme and description of the feature vectors used in these analyses can be found in Table 3.2. In each analysis, data from two participants (one from each native-language group) were used for the test set in each fold while the remaining participants' data were used for training. This led to an eleven-fold cross-validation procedure for each of the analyses.

### Statistical Analyses

The significance levels of individual participant's classification results in both the within- and cross-participants analyses were determined based on the estimated binomial confidence intervals for the number of data epochs available (Müller-Putz et al., 2008). The same procedure was used to evaluate the results of the native-language decoding analysis. Two-way repeated-measures ANOVAs with either stimulus condition (factor levels: 63 ms VOT, 41 ms VOT and 19 ms VOT) or data set (factor levels: 'individual', 'Cross-PP Native' and 'Cross-PP All') as within-subjects factor and native language (factor levels: 'English' and 'Dutch') as a between-subjects factor were used to determine whether these variables influenced classifier performance. Subsequent within- and between-subjects comparisons were carried out using paired-samples and independent-samples t-tests, respectively.

## Results

A summary of the behavioral results from Chapter Two for the stimulus conditions analyzed in the present study are presented in Figure 3.2a. Grand averaged ERP responses to

| Analysis | Feature Vector | Description |
|---|---|---|
| Single-Trial | 64 x 540 | Concatenated single-trial ERPs for standard and deviant trials in all three measurement conditions |
| Single-Trial (Combined) | 64 x 540 | Combined single-trial predictions (70) per participant |
| Grand-Average A | 64 x 540 | Concatenated individual grand-average ERPs for standard and deviant trials in all three measurement conditions |
| Grand-Average B | 64 x 180 | Concatenated individual grand-average ERPs for 63 ms VOT deviant and 41 ms VOT deviant |
| Behavioral | 1 x 7 | Mean individual behavioral responses to the stimulus continuum used in the original study |

Table 3.2: *Features used in decoding analyses of native language groups. Feature vectors are described in terms of [channels] x [time points], with the exception of the behavioral analysis, which included mean individual responses to each of the 7 stimuli in the continuum used in Chapter Two.*

the standard and deviant stimuli across the three measurement conditions for both the native-English and native-Dutch groups are presented in Figure 3.2b, along with difference waves obtained by subtracting the grand average ERP for the standard stimulus from that of the deviant stimulus in each condition. AUC scores for spatio-temporal features in the analyzed data are presented for native-English and native-Dutch speakers in each of the three measurement conditions in Figure 3.2c. ERPs collected for deviant stimuli were primarily characterized by enhancements of three negative components relative to the ERPs collected for the standard stimuli immediately preceding them: the N1, the N2 (the time interval where MMN analysis was performed in Chapter Two) and a late negativity corresponding to the RON (Schroger and Wolff, 1998; Munka and Berti, 2006). The relative difference in amplitude of these three components is most easily seen in the difference wave plots in Figure 3.2b. These same time points are also visible the AUC scores plotted in Figure 3.2c. Generally speaking, the differences in the response amplitudes of these components in the standard and deviant ERPs increased as a function of the distance in VOT between the standard and deviant stimuli, with differences being the largest in the most deviant (19 ms VOT) condition.

## Within-participant classification of phoneme contrasts

The results of the within-participant analyses, along with group means and significance levels for individual results, are presented in Figure 3.3a. A significant main effect of stimulus condition (63, 41 or 19 ms VOT deviant) was found ($F(2, 40) = 16.894, p < .0001$), along with a marginal effect of native language group ($F(1, 20) = 3.993, p = .06$). On average, classification rates increased as the difference in VOT between the standard /pa/ and the deviant stimulus grew larger, with classification rates for the 19 ms VOT deviant (across-category) significantly higher than those of both the 63 ms VOT deviant (within-category) ($t_{21} = -6.552, p < .0001$) and the 41 ms VOT deviant (category-

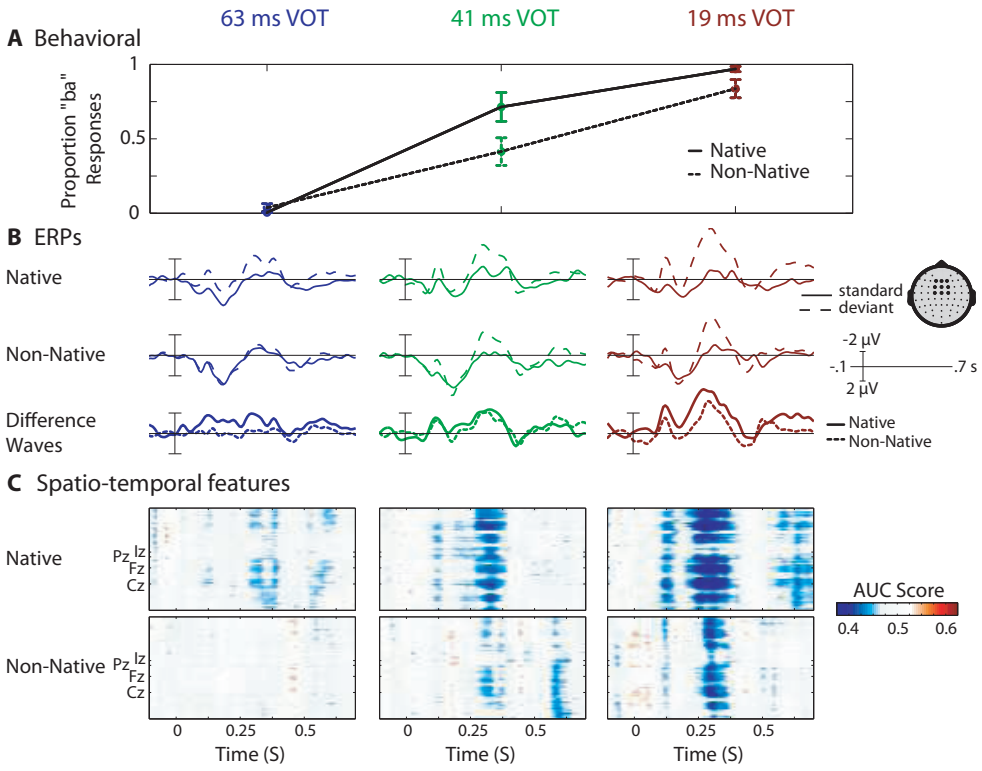


Figure 3.2: *Group level behavioral and ERP responses. A) Mean behavioral identification scores for native and non-native speakers for the three deviant stimuli. B) Group-level ERPs for both the standard and deviant stimuli are presented in each of the three measurement conditions for both native-English and native-Dutch participants. Responses are averaged across nine fronto-central electrode locations, indicated by the large dots in the scalp map presented above (see also Table 3.1). In addition, difference waves have been derived for each language group by subtracting the grand-average responses to the standard stimulus from that of the deviant stimulus in each of the measurement conditions. C) Area under the ROC-curve scores for spatio-temporal features across the three deviant conditions for both native and non-native participants. The relative locations of four midline electrodes are indicated for reference.*

boundary) ($t_{21} = -3.695, p = .001$). Additionally, mean single-trial classification rates in each of the three analyses were higher overall for the native-English speakers than for the native-Dutch speakers, with the difference reaching significance for the 63 ms VOT deviant ($t_{20} = 2.71, p < .05$).

Figure 3.3b plots the relationship between individual classifier performance across different conditions and the mean individual MMN amplitudes measured in Chapter Two at fronto-central locations (see Table 3.1) in the same conditions ($r = -.31, p < .05$), with more negative mean amplitudes tending to correspond with higher classification rates. Figure 3.3c plots the relationship between individual mean classifier decision rates obtained per stimulus when training a classifier using data from all three conditions (standard /pa/ and the three deviant stimuli) and the individual behavioral identification scores from Chapter Two for the same stimuli. A strong relationship between the classifier decision rates and the individual identification rates was found ($r = .64, p < .0001$), with stimuli classified as deviants more likely to be identified as /ba/ by participants.

An additional analysis of the classifier predictions obtained in the 19 ms VOT condition was performed to determine the performance benefits obtained when combining classifier predictions from multiple successive data epochs. These results are plotted in Figure 3.4. As one would expect, classification rates increased on average with each additional trial of data that was included. Moreover, the benefit gained from an increased number of trials was related to the single-trial classification rate. Participants with high single-trial classification rates reached rates above 0.9 when using 7 trials of data, while participants with low single-trial rates showed relatively little improvement and even a drop in performance.

## Cross-participant classification of phoneme contrasts

Cross-participant classification results are plotted in Figure 3.5, along with the individual within-participant results for the same condition (19 ms VOT deviant, across-category). Significant main effects of data set ('individual', 'Cross-PP Native' or 'Cross-PP All', $F(2, 40) = 4.688, p < .05$) and native language ($F(1, 20) = 11.995, p < .01$) were found along with a significant interaction of the two variables ($F(2, 40) = 9.084, p = .001$). On average, classifier performance was significantly higher when trained using data from all participants than when trained using individual participant's data sets ($t_{21} = -3.768, p = .001$). This is possibly due to the fact that a larger number of examples were used in training this cross-participant classifier. It might also be the case that the nature of the between-participant variability reflects non-essential sources of information, which in turn help prevent the classifier from over-fitting the training set data in the individual folds. No significant difference was found in classifier performance when trained on the 'Cross-PP Native' dataset as compared to either the within-participant classifier performance nor the classifier trained on 'Cross-PP All' dataset. However, when comparing the mean rates of the two groups across data sets, classifier performance was significantly higher for native-English speakers when trained using the 'Cross-PP Native' dataset ($t_{20} = 4.626, p < .0001$), and marginally so when trained on the 'Cross-PP All' dataset ($t_{20} = 2.073, p = .05$).

A singular value decomposition of the weight matrix of the classifier trained on the 'Cross-PP All' dataset was performed to identify the topography and time course of the components which explain the largest portion of the classifier's overall performance. The largest of these components is plotted in Figure 3.6. As can be seen, this component

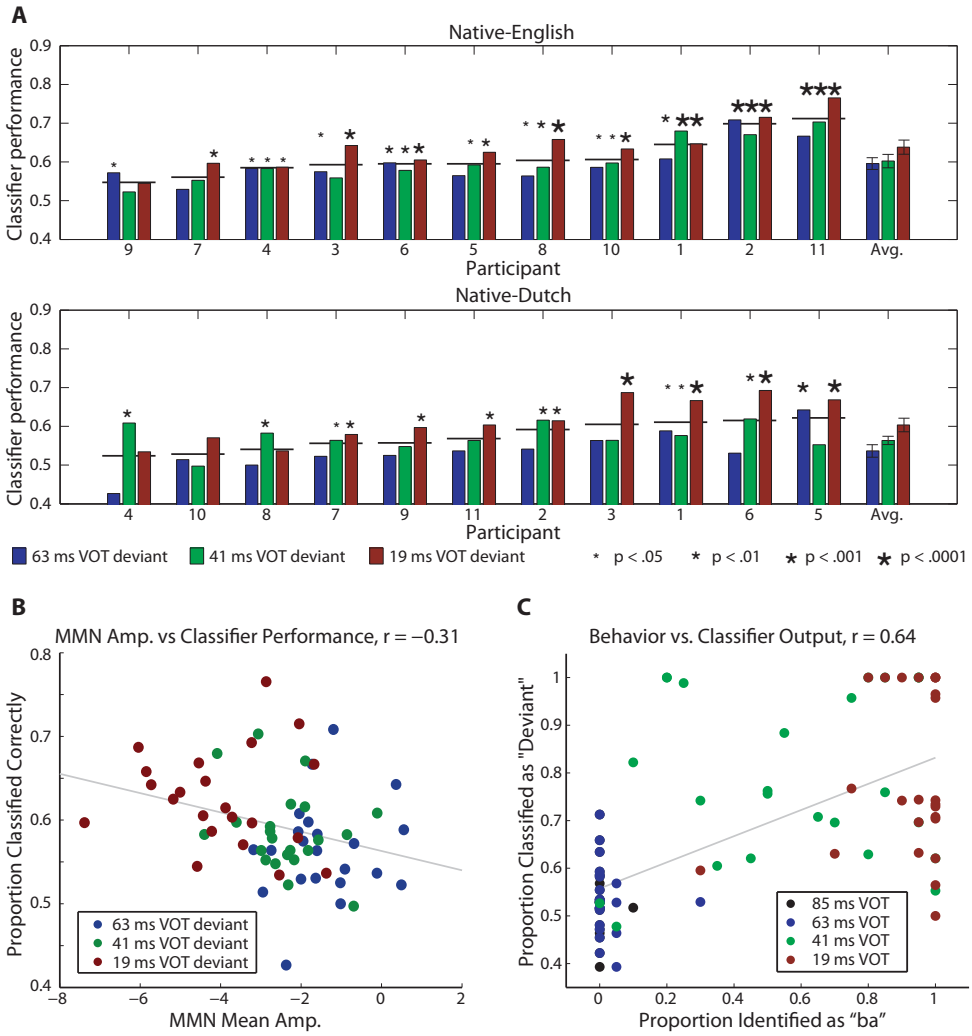


Figure 3.3: *Within-participant classification analyses. A) Classification rates for native and non-native participants for each of the three stimulus conditions along with group averages (shown with error bars). Participants are sorted based on the averaged results of the three analyses, as indicated by the horizontal lines. Asterisk size indicates the significance level of the result in each of the three conditions. B) Scatter plot of classifier performance with respect to the mean amplitude of the MMN component of individual ERPs measured in Chapter Two. C) Scatter plot of mean classifier decision rates per condition with respect to behavioral decisions in the identification task reported in Chapter Two.*

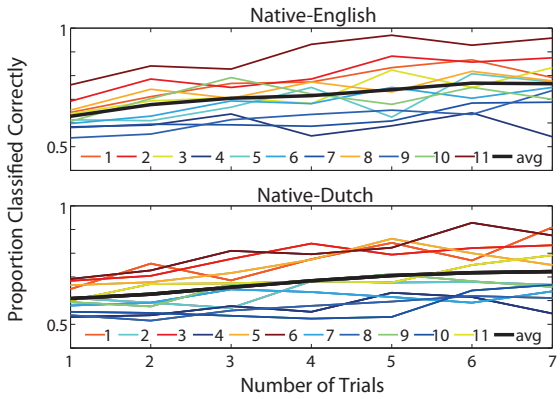


Figure 3.4: *Classification across multiple trials for the 19 ms VOT condition. Multi-trial performance for individual participants in both the native-English and native-Dutch participant groups is shown using colored lines (sorted according to mean individual performance), while the average for each group is shown using a thick black line. On average, performance increased when including additional trials. Participants with relatively high single-trial classification rates tended to show additional improvement when decisions were based on additional trials, while participants with low single-trial classification rates showed less benefit from the inclusion of additional trials.*

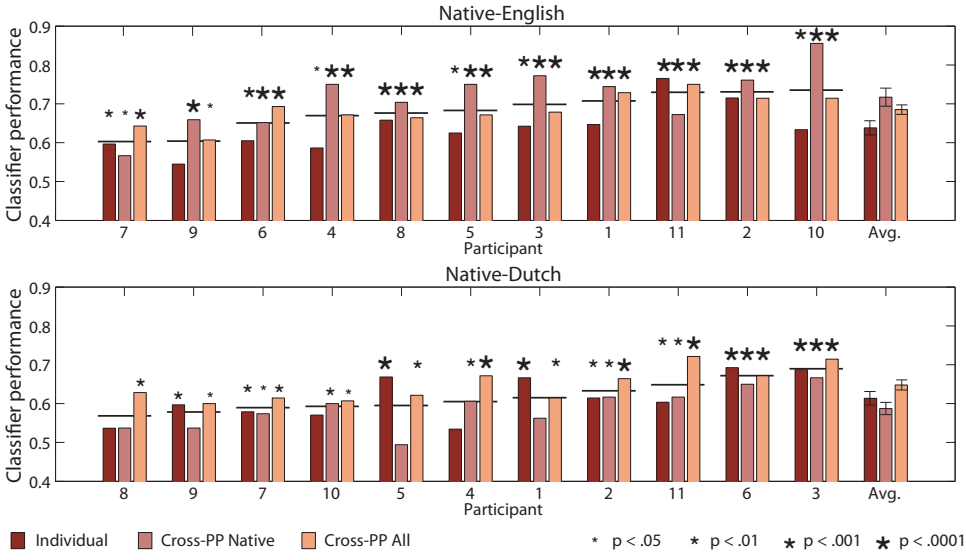


Figure 3.5: *Cross-participant classification analyses. Classification rates for native and non-native participants for the two classifiers trained on cross-participant data sets using the 19 ms VOT deviant, along with individual rates from the within-participant classification analysis of the same deviant condition. Results for each of the three datasets are indicated using different colored bars. Participants are sorted based on the averaged results of the three analyses, as indicated by the horizontal lines. Group averages are also shown with error bars. Asterisk size indicates the significance level of a given individual result.*

explains about 44% of the variance in the classifier weighting matrix, has a negative fronto-central distribution typical of the MMN response (Duncan et al., 2009), and which highly resembles the difference wave time courses during the peak of the ERP responses for the 19 ms VOT deviant condition presented in Figure 1. Moreover, a high correlation ($r = -.946, p < .0001$) between this component's time course and the average of the difference waves of the ERPs for all participants at the same time points indicates a strong relationship between the classifier weighting matrix and the ERPs.

### Decoding of native language

The mean results for each of the five analyses are plotted in Figure 3.7. Classifier performance was significantly above chance for four of the five data sets which were analyzed: single-trial data, individual average ERP data, individual average ERP data for 63 and 41 ms VOT deviants, and individual behavioral data. The exception was when using the combined single-trial predictions. The highest overall rate of 83% was attained when using only averaged individual ERP data from the 63 ms and 41 ms VOT deviant stimuli. These were the two conditions which showed a significant between-groups difference in the ERP analysis in Chapter Two. Classifier performance was slightly lower when using the ERP data from all measurement conditions, followed by the analysis in which the vector of mean individual identification scores was used.

## Discussion

The present study investigated the outcomes of a series of multivariate pattern classification analyses of EEG data collected during passive speech perception of English phonemes by native and non-native listeners. These analyses addressed two principal research questions: 1) Is it possible to decode stimulus categories from single-trial EEG data elicited using different speech sound contrasts for native-English and native-Dutch speakers? 2) Is it possible to decode these same stimulus categories across individual participants, either within or between native-langauge groups?

### Within-participant analyses

The results of the within-participant analyses demonstrate that single-trial EEG measurements of brain responses to phonemes contain sufficient information to decode speech sound categorization, and that the performance of such analyses improved across conditions representing increasingly salient phonetic contrasts. As such, the results confirm that the within-participants trends previously observed in Chapter Two's analyses of grand-averaged ERP data are also present at the single-trial level. In the case of the 19 ms VOT condition, which employed a stimulus contrast that clearly represented two distinct phonetic categories for both native-English and native-Dutch participants, classifier performance was significantly higher than conditions which employed a within-category (63 ms VOT) or ambiguous (41 ms VOT) contrast. Previous research findings have shown enhancements of different components of the auditory ERP to across-category deviant stimuli as compared to within-category deviant stimuli, including the MMN (Sharma and Dorman, 1999) and the N2/P3 complex (Maiste et al., 1995). Enhancements of these components

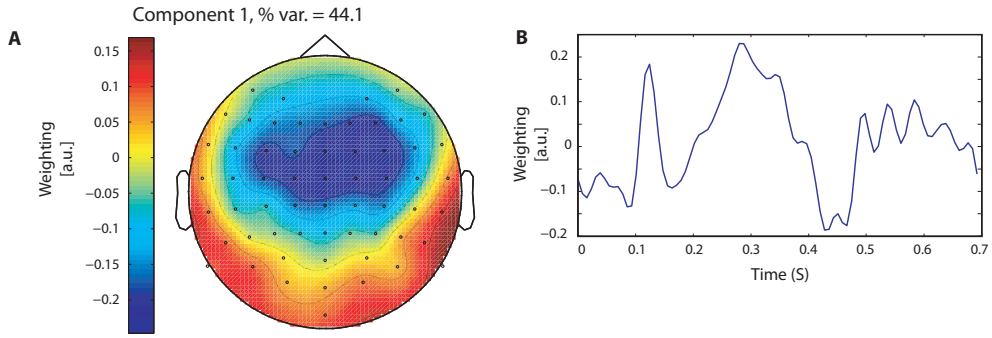


Figure 3.6: *A) Topography and B) time-course of first component obtained through a singular value decomposition of the classifier weights trained on data from all 22 participants. The data are presented in an arbitrary unit scaling.*

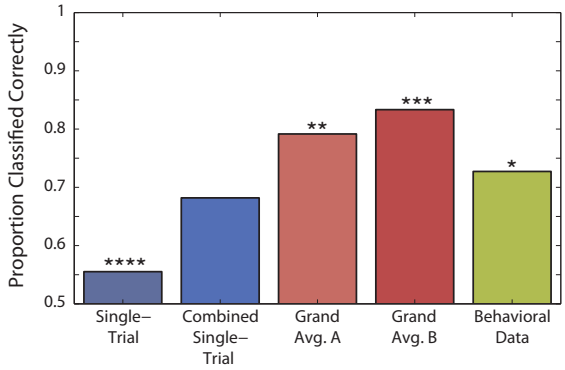


Figure 3.7: *Prediction of native language on the basis of electrophysiological and behavioral data. Five decoding analyses aimed at predicting the native-language of a given-participant on the basis of their measured data were carried out. Two analyses made use of concatenated single-trial EEG data from each of the three measurement conditions. The first of these analyses determined single-trial classification rates using this data set, while the second combined the single-trial predictions (70 total trials) for each participant's data obtained when it was used as a test-set during the classification analysis. Two additional analyses made used of concatenated individual grand-averaged ERPs. One utilized both standard and deviant stimulus ERPs collected in all of the three measurement conditions, while the other included only the deviant ERPs measured using the 63 and 41 ms VOT stimuli. A final analysis was performed using a vector of seven mean behavioral identification scores collected for each participant in Chapter Two. Significance levels are shown using asterisks ($p < .05 = *$, $p < .01 = **$, $p < .001 = ***$, $p < .0001 = ****$), and are based on the number of observations available for each of the five data sets. For the single-trial analysis, 1540 data points (70 per participant) were available, while for the remaining four analyses, 22 data points (one per participant) were available.*

in deviant trials would in principal increase the amount of information available during pattern classification, leading to higher overall performance.

A marginally significant effect of native language was also observed, suggesting an overall difference in decoding performance across the two groups. In general, decoding rates were higher for the native-English speakers as compared with the native-Dutch speakers, with this difference reaching significance for the 63 ms VOT condition. When looking at the AUC scores presented in Figure 3.2c, we see a clear difference in the amount of discriminative information available between the two language groups, not only in the 63 ms VOT condition, but across all three conditions. This could in part explain the overall differences which were observed in decoding performance between the two groups. It is worth noting again that the native-Dutch group of participants were in fact highly proficient English speakers, having undergone 6 years of coursework as part of their high school curriculum. The fact that differences in decoding performance are still observed highlights the formative role played by language learning in early childhood in shaping our long-term perception of speech (Werker and Tees, 1984; Best et al., 1988; Werker and Lalonde, 1988; Kuhl et al., 1992; Best, 1993; Kuhl, 2004).

A modest correlation was observed when investigating the relationship between individual decoding results and individual mean MMN amplitudes measured in the original analysis in Chapter Two. The fact that this relationship was not stronger may be because the multivariate analysis also included time points and scalp locations which were not part of the original analysis. While the individual MMN amplitudes reflect the activity in a 50 ms time window around the peak of the MMN difference wave at fronto-central locations between 200-400 ms following stimulus presentation, the classification analysis included data from all 64 recording channels at time points between 0 and 700 ms post-stimulus onset. As such, it included additional ERP components including the N1, P3a and RON. Previous work on single-trial classification of ERP components has shown that the inclusion of components at different time intervals within an ERP provides additional information when distinguishing different classes of signals, leading to an improvement in classifier performance (Blankertz et al., 2011). So while a significant relationship was observed between the MMN component of individual ERPs and the results of the within-participant classification analyses, it would appear that decoding performance is also influenced by a broader hierarchy of cognitive processes underlying the responses observed at different time points.

A much stronger correlation was found when examining the relationship between the individual behavioral data collected in Chapter Two and the per stimulus classification performance observed when training a classifier using data from all three conditions. This seems to suggest an overlap in the functional organization underlying both the perceptual decision making process during behavioral identification and the single-trial brain responses used by the classifier during its decision-making process. This result is perhaps most interesting when we consider the fact the behavioral identification measurements reflect an active process (responding to individual stimuli) while the EEG measurements reflect a passive process (perception of sound sequences while viewing a film).

Within-participant single-trial classification rates were comparable with the average rates reported in Herrmann et al. (2012). The results of that study also showed a graded pattern of results depending on which manipulation (auditory space, syntax or both) was present in the experimental stimulus. Here, the graded responses are observed relative to a continuous change in one specific acoustic feature of the deviant stimuli (VOT), as

well as with respect to the native language of individual participants. When compared to the average single-trial rates observed in experiments making use of an active auditory listening task (Halder et al., 2010; Schreuder et al., 2010; Höhne, 2011; Vlek et al., 2011), the rates reported here are substantially lower. This is most likely due to the fact that the tasks in the studies just mentioned were designed to elicit the P300 response, which has a substantially higher amplitude (10-20 $\mu V$) than the ERP components elicited during passive listening, such as the MMN (0.5-5 $\mu V$) (Duncan et al., 2009; Gonsalvez and Polich, 2002). Such increases in signal amplitude lead to a higher signal-to-noise ratio, and improve classification performance.

It was also shown that this performance could be improved through the inclusion of additional trials. Performance increased on average with each additional trial that was included, reaching above 95% correct for some participants when 7 trials were included in the classifier's decision. However, the relative benefit in classification performance which was achieved through the use of additional trials was also a function of individual participant's single-trial classification rates. While individuals with relatively good single-trial classification rates tended to show the most improvement across trials, participants with low single-trial classification rates did not show much benefit when including additional trials, with performance sometimes being even lower than the single-trial rates. This would seem to point to a general lack of discriminative information in the single-trial EEG data for some participants. Previous multivariate pattern classification analyses of EEG-data collected in an auditory paradigm and using multiple-trials have also shown similar results (Schaefer et al., 2011). Such differences may be due in part to what has been referred to as 'BCI illiteracy', in which some participants do not show a neural signature of interest for a given task (Vidaurre and Blankertz, 2010). Previous studies on individual MMN responses have also demonstrated that not everyone will show a clear MMN component despite exhibiting normal auditory perceptual abilities (Bishop and Hardiman, 2010).

## Cross-participant analyses

One of the goals of the present analysis was to determine the amount of individual overlap in the functional brain organization underlying the perception of the phonemes used during EEG measurements, both within and across native-language groups. When using a classifier trained on data from 10 of the 11 native-English participants collected in the 19 ms VOT condition ('Cross-PP Native'), a difference in the classifier's performance was observed for the two language groups. While performance improved for native-English speakers relative to the within-participant analysis (64% vs 72% correct), performance decreased for native-Dutch speakers (61% vs 59%). In contrast, when using a classifier trained using data collected in the same condition from all 22 participants ('Cross-PP All'), a significant overall improvement was observed for all participants relative to the within-participant analysis. Here the performance benefit for native-English speakers was slightly less as compared with the benefit seen when using a classifier trained using only data from native-English speakers. This seems to indicate a discrepancy in the extent to which features present in the single-trial data of native-English speakers are utilized by the two cross-participant classifiers, and that features present in the single-trial data of the native-Dutch speakers do not completely overlap with those of the native-English speakers.

Previous work using fMRI to investigate differences in the functional neuroanatomy of language processing between native and non-native speakers suggests that, while both

groups rely on the same cortical network, non-native speakers show enhanced activation in some regions relative to native speakers (Rüschemeyer et al., 2006). Studies using ERP measurements have suggested an enhancement of ERP components related to the processing of both acoustic features (see Chapter Two) and categorical information (Näätänen et al., 1997; Winkler et al., 1999) measured with native speakers relative to non-native speakers. Combined, these results suggest both similarities in the functional organization of language processing in native and non-native speakers as well as differences in the distributed activation patterns for specific linguistic tasks. The present cross-participant analyses provide additional support for this view. They are also in line with previous cross-participant classification analyses presented in Schaefer et al. (2011) and Herrmann et al. (2012), which showed either equivalent or improved classification performance when using cross-participant data sets as compared to within-participant datasets. As was the case with the results presented in Schaefer et al. (2011), the overall improvement in performance here may be due to the increased amount of training data available in the cross-participant analysis.

This study also presented the results of a set of cross-participant classification analyses that focused on the native language of participants. Analyses that made use of single-trial ERP data were less successful at determining the native language of a given participant than those which made use of individual behavioral data. However, analyses which made use of individual grand-averaged ERPs showed better native-language classification than the analysis using behavioral data, with the best overall performance obtained when using ERPs measured in response to the 63 ms VOT and 41 ms VOT deviant stimuli. These were the two conditions which showed a significant between-groups difference in MMN response amplitude in Chapter Two. These results suggest that our brain responses to speech may reveal more about our linguistic background than our behavioral responses to it. They also align nicely with the results of the cross-participant analyses discussed above, in that they also suggest differences in the distribution of activation patterns measured in response to speech stimuli between native-English and native-Dutch speakers.

## BCI paradigms based on speech perception

The use of multivariate pattern classification methods to identify differences in the characteristic brain responses generated by individual members of groups with differing perceptual profiles could have potential applications in both education and clinical settings. A new class of BCIs has recently been described, called passive BCIs, which combine cognitive monitoring with the real-time decoding methods typical of BCIs (Zander and Kothe, 2011). A passive BCI based on the listening paradigm used in this study could be used to monitor the brain activity underlying auditory perception. In educational settings, such a system could be used to ascertain whether one's brain responses to foreign speech sound contrasts resemble those of a native speaker or not. Likewise, in clinical settings, characteristic abnormalities in the MMN component have been reported for a wide-variety of clinical populations, including children with specific language impairment and individuals diagnosed with schizophrenia (Näätänen et al., 2007). In turn, the use of an appropriate BCI may be able to reduce the measurement times which are needed in order to ascertain whether an individual's brain responses fit a particular neurological profile. However, some caution is needed when considering such approaches. Many ethical issues arise when considering the applications made possible by single-trial decoding approaches, including the

unwilling extraction of personal information from measurements of brain-activity and their potential (mis)use in criminal investigations (Haynes and Rees, 2006; van Gerven et al., 2009; Haselager et al., 2009).

The present results also suggest that BCIs which directly support language learning through neurofeedback have potential. Neurofeedback presents users with real-time information about brain activity as measured using EEG or fMRI, providing them a mechanism to modulate activity related to specific brain structures or cognitive states (Yoo et al., 2006; Johnston et al., 2010; Zoefel et al., 2011; Zotev et al., 2011; Weiskopf, 2012). In a recent study (Shibata et al., 2011), multivariate methods were employed in conjunction with fMRI measurements of activity in striate and extrastriate cortical regions during visual perception of simple orientation stimuli, and were subsequently used to provide participants with a neurofeedback signal based on decoded brain activity from these same regions. Following 5-10 days of neurofeedback training, participants showed enhanced visual perception of stimuli corresponding to the trained activation patterns.

This type of induced perceptual learning may also be possible using decoded-EEG neurofeedback based on the evoked responses underlying speech perception. Such a system would, in principle, provide users with real-time information regarding their brain's ongoing responses to unfamiliar foreign speech sound contrasts, as reflected in the MMN and other components of the auditory evoked response. Research on the time course of language learning and associated changes in brain responses has shown that the MMN response develops prior to changes in behavioral responses associated with the successful discrimination of foreign phoneme contrasts (Tremblay et al., 1998). Thus it would be possible to provide users with neurofeedback in a time span where the perceptual learning process is still ongoing. The results of the multi-trial analysis presented above also suggest that the reliability of such feedback could be regulated by combining classifier decisions across a sufficient number of subsequent trials. Moreover, it may also be possible to make use of classifiers trained on cross-participant data sets from, for instance, native speakers, for use in neurofeedback paradigms intended for second-language learners. While additional research would obviously be needed to verify the merit of this approach, the results presented here in conjunction with those from Shibata et al. (2011) suggest that such an approach is possible. Many challenges remain in the development of such a system. For example, it is still an empirical question how high the single-trial classification rate has to be to support language learning. Nevertheless, the above-chance single-trial classification reported here is promising. It indicates that, at least with respect to the multivariate pattern classification that would be required, a neurofeedback system for the training of speech perception is feasible.

# Conclusion

The present study has shown that both within- and cross-participant decoding of evoked responses measured during speech perception is possible, with the results being a function of both the relative size of the contrasts employed as well as the phonological status of the contrast for a given listener. Moreover, the results indicate that, while the functional brain organization underlying speech perception may involve the same fundamental networks in native and non-native speakers, differences in the relative distribution of activation patterns influence the outcomes of the multivariate analyses for native and non-native speakers. On

the basis of these results, we suggest that these methods can be used for developing novel BCI applications related to second language learning.

# 4

# Modulation of auditory evoked responses using online decoded-EEG neurofeedback: Towards enhanced perceptual learning

*based on Brandmeyer, A., Sadakata, M., Spyrou, L., McQueen, J.M. & Desain, P., 2014. Modulation of auditory evoked responses using online decoded-EEG neurofeedback: Towards enhanced perceptual learning. In preparation.*

**Abstract:** One application of the BCI paradigm is to provide feedback to users on different mental states that can be decoded at the single-trial level using classification methods. This study investigated the effects of such a feedback paradigm on behavior (frequency discrimination sensitivity) and auditory evoked potentials (AEPs) using a passive listening task and naive participants. Participants completed four sessions of EEG measurements, each of which included neurofeedback. During feedback, AEPs elicited using oddball sequences and containing the mismatch negativity (MMN) and P3a responses were analyzed using a logistic regression classifier trained on individual participants' data. This analysis served as the basis for a feedback signal, which controlled the clarity of films being viewed by participants. A matched control group underwent an identical procedure, with the exception that they received sham feedback based on the brain responses of the experimental group, rather than their own. No effect of the feedback on behavior was observed. Analysis of AEPs collected during feedback revealed that MMN responses of the experimental group were enhanced relative to those of the control group. These findings indicate the potential for decoded-EEG neurofeedback to modulate targeted forms of brain activity. Such feedback has applications in domains where reductions of specific forms of brain activity are associated with functional deficits, for instance, in educational or clinical settings.

# Introduction

While neurofeedback research and applications have been proceeding for over 40 years, they have predominantly focused on the use of EEG activity in specific frequency bands of interest, such as alpha and theta, due to the ease with which these features can be extracted from the data (Hammond, 2011). Only recently has there been an interest in the use of other types of neural markers and the use of multivariate pattern classification methods in neurofeedback paradigms (Pham et al., 2005; Shibata et al., 2011). One of the most exciting potential applications of such neurofeedback paradigms is in the domain of perceptual learning. A study using fMRI and neurofeedback training based on the decoding of activity in early visual cortex regions involved in the perception of orientation stimuli recently demonstrated improvements in visual perception (Shibata et al., 2011). Here, the results of a study which investigated a similar decoding-based approach using EEG, in the domain of auditory perceptual learning, are presented.

EEG has been used extensively to study the effects of perceptual learning on different components of the event-related potential (Atienza, 2002; Reinke et al., 2003; Alain et al., 2006). In particular, the mismatch negativity (MMN) component has been shown to reflect individual differences in the perception of simple features such as the frequency of a sound (Koelsch et al., 1999) or more abstract features such as phonetic categories (Näätänen et al., 1997; Winkler et al., 1999). The MMN is typically observed as a frontocentral negativity between 100-300 ms following the presentation of an oddball stimulus during passive listening (Näätänen et al., 2007; Kujala et al., 2007; Duncan et al., 2009). Several studies investigating the time-course of perceptual learning have shown a progressive enhancement of the response as participants become more proficient at perceptual discrimination (Tremblay et al., 1998; Menning et al., 2000). A neurofeedback paradigm based on the enhancement of the MMN response could therefore possibly be used to enhance perceptual learning.

Additional support is provided by studies investigating the application of multivariate pattern classification to datasets collected using an MMN paradigm (Herrmann et al., 2012; Tzovara et al., 2013; Brandmeyer et al., 2013). Studies reporting on both the perception of simple acoustic deviations (Herrmann et al., 2012; Tzovara et al., 2013) as well as linguistic features (Herrmann et al., 2012; Brandmeyer et al., 2013) have reported average single-trial classification rates between 60-70% (standard-deviant), which can also be improved through the use of multiple trials (Brandmeyer et al., 2013). While feedback based on such classification rates is noisy, due to occasional incorrect feedback, it can still lead to learning effects (Cardinal, 2006).

To investigate the effects of a decoded-EEG neurofeedback paradigm on the MMN response and perceptual learning, an experiment was designed based on the perception of frequency differences in sinusoidal tones. These stimuli have been widely used both in MMN research (Sams et al., 1985; Näätänen et al., 2007; Duncan et al., 2009) and psychoacoustic experiments investigating auditory perceptual learning (Micheyl et al., 2006). The characteristics of the event-related potentials they elicit and the time-course and parameters of auditory perceptual learning related to them are well understood.

Neurofeedback was provided in the form of changes to a visual signal. During the sessions, incoming EEG data containing the MMN was classified, and the output of this classifier was used to control a blurring filter applied to silent movies. Other neurofeedback paradigms have also made use of visual reward signals (Hammond, 2011; Shibata et al.,

2011). Such an approach also takes advantage of a common feature of MMN measurements, which is the viewing of silent films. This is done because the MMN response is pre-attentive in nature, and provides participants with an easy task to perform (Näätänen et al., 2007).

The participants were unaware of the neurofeedback procedure, and instead completed a visual-perception decoy task. This was done because the MMN is a pre-attentive response, and because a control group was provided with sham feedback. More specifically, it was done in order to remove the possibility that differences in awareness of the manipulation would lead to differences in attention or motivation between the two groups. Rather, participants were instructed to concentrate on the films in order to answer questions about their content, and that the blurring filter and sound sequences would act as distractions and make answering the questions more difficult.

There were two key research questions. Firstly, does the presentation of this form of decoded-neurofeedback lead to measureable differences in the ongoing brain responses of the two groups of participants? This point was investigated by analyzing the ERPs (MMN and P3a components) of the two groups prior to and then during neurofeedback. Secondly, if any changes are observed, do they correspond to any differences in the two groups' perceptual discrimination abilities? This was assessed by having participants complete a listening task following each session. Enhancement of performance in the group receiving neurofeedback would suggest a beneficial effect on perceptual learning; specifically, enhanced frequency discrimination sensitivity.

# Methods

## Participants

16 participants with normal hearing and normal or corrected to normal vision completed four days of measurements. Participants were recruited using the Radboud University experiment-participation system. All participants were students at the university, and provided written informed consent. Participants were paid 10 euros per hour. They were divided evenly into the 'experimental group' (7 females, mean age = 22.4 years) and the 'control group' (6 females, mean age = 22.3 years).

## Procedure

The design is presented in Figure 4.1. The experiment consisted of four sessions on separate days within one week. All measurements took place in an acoustically attenuated, electrically shielded cabin. On the first day, participants first completed up to three repetitions of a listening task to determine their frequency discrimination threshold. Afterwards they completed an initial EEG measurement session in which MMN responses were measured. All participants were required to show thresholds of less than 2% ($\Delta Hz$), mean MMN amplitudes more negative than -1 $\mu V$, and single-trial classification rates above 58% (see below for additional details regarding these measures). This assured that all participants were matched to a specific level of performance at the onset of the experiment, and that the neurofeedback was based on above-chance classification performance. This led to the exclusion of an additional 12 participants.

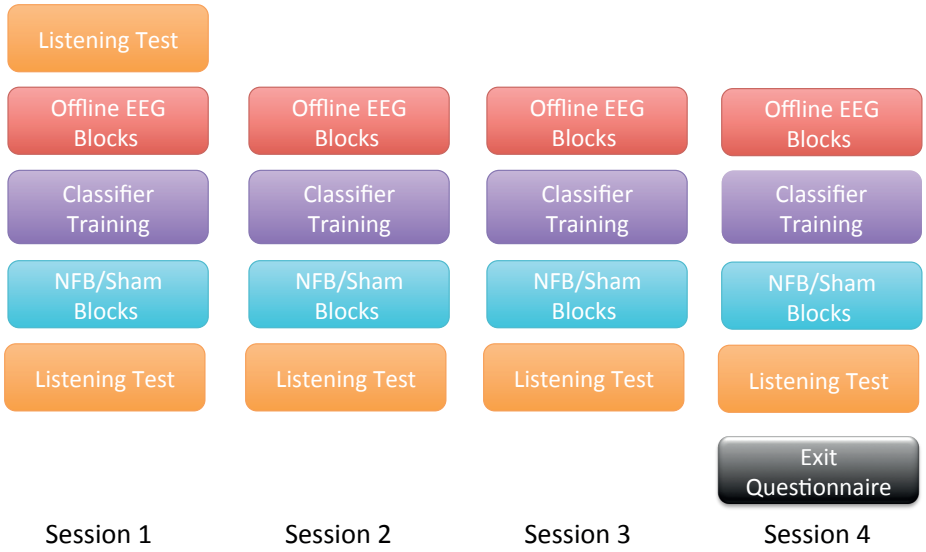


Figure 4.1: *Experimental design and procedure.*

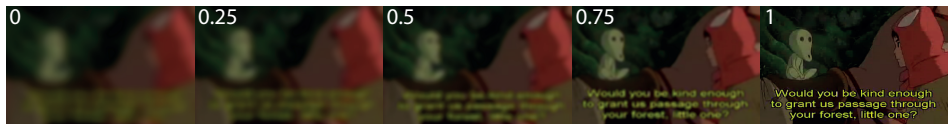


Figure 4.2: *Visual feedback used in the experiment. Each panel illustrates the blurring effect obtained using feedback values ranging from 0-1.*

## Stimuli

Auditory stimuli consisted of sinusoidal tones with fundamental frequencies ($f0$) between 500 to 550 Hz with a duration of 100 ms. They were presented over Etymotic ER-4P insert headphones at approximately 70 dB SPL. All stimuli were rendered as 16-bit digital audio at a sample rate of 44.1 kHz.

During the EEG measurements, participants viewed silent films with subtitles on a 15.4 inch TFT screen with a resolution of 800 x 600 pixels. Films were presented using a smaller area (600 x 400 pixels) centered within the screen to reduce eye movements. Participants were seated such that the distance between their eyes and the screen was approximately 60-70 cm.

## Listening Test

A 2AFC staircase procedure (3-up 1-down) was used to estimate the individual participant's 79% correct discrimination thresholds (Levitt, 1971). In each trial, two tones were presented with an interstimulus interval of 1 second. One tone always had an $f0$ of 500 Hz, while the other tone's $f0$ was adaptively varied downwards, starting at 525 Hz (i.e 5% $\Delta Hz$). The order of the two tones was randomized. Participants were asked to judge which tone came first. The step size was varied in the first 4 reversals, using steps of 10, 5, 2.5 and ending with 1 Hz. Each measurement terminated following either 16 reversals or 100 trials.

## EEG Measurements

EEG was measured using a 64-channel BioSemi ActiveTwo amplifier along with left/right mastoid leads, and horizontal and vertical EOG at a sample rate of 256 Hz using active AgCl electrodes. Electrode placement was performed in accordance to the international 10-20 system.

The EEG measurement blocks were conducted while participants viewed silent animation films with subtitles and passively listened to oddball sequences of tone stimuli. The oddball sequences contained 85% standard stimuli and 15% deviant stimuli, with 3 to 9 standards occurring between deviants and at the beginning of each block. Each block contained 1000 total trials and lasted approximately 8.3 minutes. An stimulus onset asynchrony of 500 ms was used.

Participants completed four blocks of offline measurements per session, which provided the data for classifier training. In two of these blocks, the $f0$ of the standard and deviant stimuli were 500 and 550 Hz, respectively, while in the other two these roles were switched. These blocks were alternated, as was their order, across participants and sessions.

During the feedback portion of each session, four additional blocks of EEG measurements using the same sequences were completed. Here, the $f0$s of the standard and deviant stimuli were always 500 and 550 Hz, respectively.

After every EEG measurement block, participants completed a 10-item question list in which they indicated whether they had seen a particular item during the section of film they just watched. The purpose of these questions was to ensure that participants had directed their attention to the films during the measurements.

Two types of trials were used during ERP analyses, classifier training and feedback: deviant trials and the standard trials immediately preceding them. This was done for two

reasons. Firstly, the novelty response to standard stimuli is known to decrease across standard trials within a given run. This means that the selected trials would contain the most diminished responses relative to the deviant trials. Secondly, the classification analyses benefit from the use of a balanced number of trials in the context of a binary classification problem.

Data were processed in non-overlapping epochs from -50-450 ms relative to stimulus onset. Electrodes with offsets greater than $\pm 35\ mV$ or with 50 Hz power exceeding 1000 $\mu V^2$ were repaired using a spherical spline interpolation procedure on a per epoch basis (Perrin et al., 1989). Data epochs were subsequently baseline corrected to the mean of the 50 ms period preceding stimulus onset, and then band-pass filtered between 1-25 Hz. Any epochs containing activity greater than $\pm 100\ \mu V$ at any channel were excluded from the subsequent analyses.

## ERP Analyses

ERP analyses were performed on the offline and feedback session data with the data referenced to the average of the left and right mastoid electrodes. Averages made use of the data from electrodes Fz and Cz, where measurements of MMN response tend to be maximal (Duncan et al., 2009). Analyses were performed on the difference waves obtained by subtracting the ERP measured using standard trials in a given set of blocks from the corresponding deviant ERP (Duncan et al., 2009). Two components in this difference wave were quantified: the MMN and P3a. For the MMN, the latency of the maximal negative peak between 80-150 ms following stimulus onset was determined, and the mean amplitude was calculated in a 25 ms window around the peak. The P3a latency and amplitude were determined using the maximally positive peak of the difference wave between 150-250 ms and the same mean amplitude procedure.

## Classification Analyses

For classification, the epoched data were further downsampled to 32 Hz, and were referenced using a surface Laplacian method (Perrin et al., 1989). Two classification analyses were performed in each of the four sessions. In each case, a binary regularized linear logistic regression classifier (Bishop, 2009) was trained using standard and deviant trials to represent the two classes. The optimal regularization strength (i.e. classifier complexity) was found using a grid search with strengths of [.001 .01 .1 1 10 100] times the total data variance. One classifier was trained using data collected from trials containing a 500 Hz stimulus, while the other was trained using trials containing a 550 Hz stimulus. This was done incrementally across sessions, such that data from all previous offline measurements were used in subsequent sessions. Additionally, for the fourth session, the regularization strength used for each classifier training was fixed to per participant values obtained during the third session (to avoid the time required to perform the search using data from all four sessions). The reported single-trial rates are based on the results of a 10-fold cross-validation procedure in which distinct partitions of the data are used for training (90% per fold) and testing (10%).

## Feedback Sessions

During feedback blocks, the two classifiers were applied to the deviant and the immediately preceding standard trials to obtain a decision value for each of the respective data epochs. For a linear logistic regression classifier, these values are calculated in the following manner:

$$f(x_i) = w^\top x_i + b \tag{4.1}$$

where $x_i$ is the EEG trace of trial i, $w$ is a set of classifier weights and $b$ is a bias term. These decision values can also be converted to posterior probabilities:

$$p(+|x_i) = 1/(1 + e^{-f(x_i)}) \tag{4.2}$$

Furthermore, these probabilities can be combined across groups of trials using a naive-Bayes formulation:

$$p(+|x_1 \dots x_n) = \frac{\prod_1^n p(+|x_i)}{\prod_1^n p(+|x_i) + \prod_1^n (1 - p(+|x_i))} \tag{4.3}$$

which is equivalent to:

$$p(+|x_1 \dots x_n) = \frac{\prod_1^n p(+|x_i)}{\prod_1^n p(+|x_i)(1 + e^{\sum_1^n f(x_i)})} = \frac{1}{1 + e^{\sum_1^n f(x_i)}} \tag{4.4}$$

During feedback blocks, two probabilities were calculated: one for deviant trials, and one for the standard trials immediately preceding them. For both trial types, overlapping sets of consecutive trials ($n = 5$) of each type were used to calculate the updated probability after each new trial was measured. The two ensuing probabilities were averaged together with equal weighting, and this value was used to adjust the visual feedback signal immediately following each deviant trial.

The visual feedback signal was created by applying a continuous Gaussian blurring filter to the films. The filter was expressed in terms of a blurring radius (i.e. number of pixels, range 0-25), which was inversely mapped onto the feedback value obtained using the output of the classifiers. This implies that a feedback value of 1 (100% posterior probability) would have a blurring radius of 0 and a feedback value of 0 (0% posterior probability) would have a blurring radius of 25. The mapping was constructed such that the subtitles of the movies were legible with feedback values of approximately .75 and above. Examples of the blurring effect can be seen in Figure 4.2.

Following each session in the experimental group, the feedback values for all relevant trials were stored offline and subsequently used as sham feedback signals for participants in the control group, maintaining block and session order. This was organized in a counterbalanced fashion across participants. The result of this procedure was that the sum of feedback signals viewed by the two groups were in effect identical, so that both groups had the same exposure to changes in the clarity of the films being viewed, realized using the same procedure. Importantly, the feedback signals received by the control participants should be uncorrelated with their ongoing brain responses.

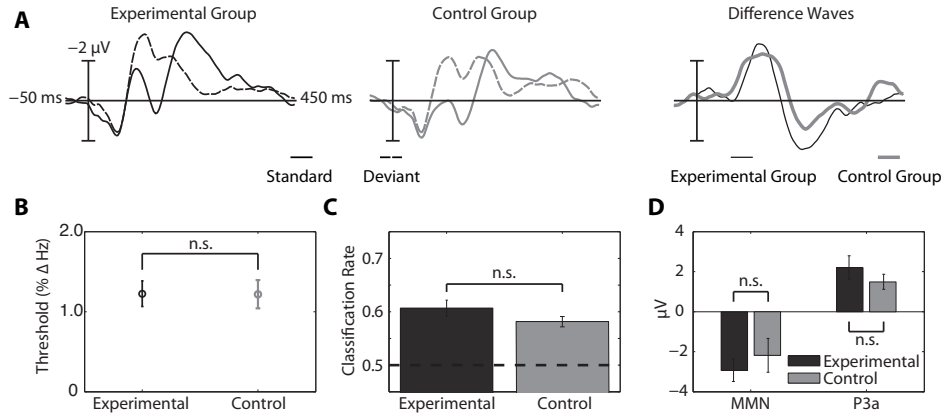


Figure 4.3: *Results of the initial offline measurements. A) Grand average ERPs and difference waves for experimental and control group participants. B) Initial behavioral discrimination thresholds. C) Single-trial classification rates for data collected during the initial offline EEG measurement blocks. Chance level (.5) is indicated with a thick dashed line. D) MMN and P3a component amplitudes. n.s. = not significant*

## Statistical Analyses

A series of one-way ANOVAs were used to investigate the effect of participant group (Experimental, Control) on the initial measurements and performance on the decoy task. Classification rates as well as data collected during the feedback sessions were analyzed using a two-way ANOVA with participant group (Experimental, Control) as between-subjects factor and session (Day 1-4) as within-subjects factor. Paired-samples t-tests were used to perform pair-wise comparisons.

# Results

## Initial Measurements

Results of the initial measurements can be seen in Figure 4.3. The two groups of participants were compared with respect to: their initial behavioral discrimination threshold, the amplitudes of the MMN and P3a response measured using the ERPs collected during the offline EEG blocks, and the performance of the classifiers trained using that data. No significant differences were found between the groups for any of the measures. Additionally, there was no significant difference in the two groups' performances on the visual-perception decoy task (Experimental Group: mean = 88% correct responses, Control Group: mean = 87%). They were therefore well matched at the onset of the experiment.

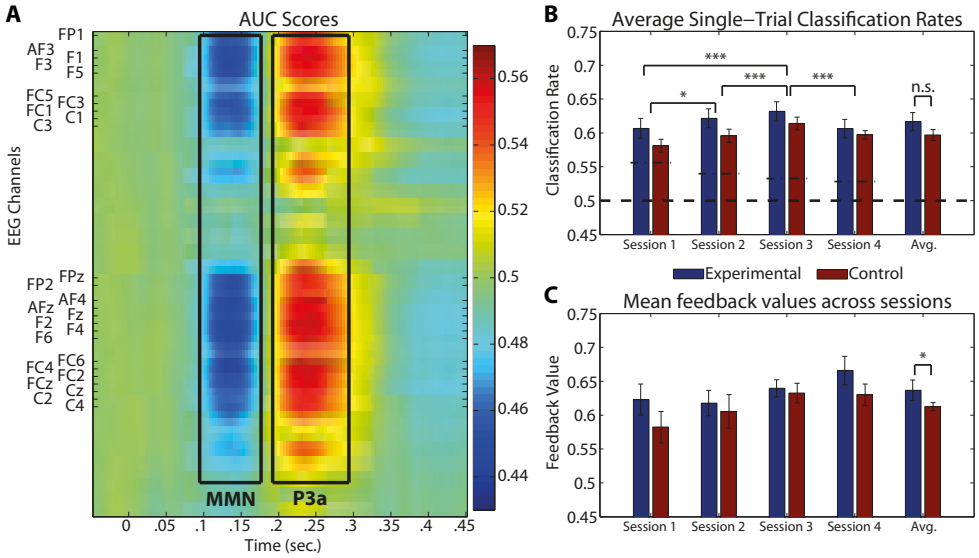


Figure 4.4: *AUC Scores, mean single-trial classification rates and feedback results. A) AUC scores. These were calculated using data from all 16 participants collected during the offline measurements over the four sessions. Two groups of features corresponding to the MMN and P3a components are indicated by the black boxes. Additionally, the 25 EEG channels with the highest mean discriminative information scores ($|AUC - .5|$) in the two time intervals are labeled. B) Mean single-trial classification rates across days. Chance level (.5) is indicated with a thick dashed line, while the binomial confidence interval for each of sessions is indicated using a thin dashed line, based on the number of available data points. Brackets are used to show the results of pair-wise comparisons based on the main effect of 'Session'. C) Feedback values across sessions. The experimental group values were those presented during the actual sessions. For the control group, feedback values were simulated post hoc using the data collected during the experiment. For B and C, statistical effects are indicated using brackets and asterisks: n.s. = not significant,* $* = p < .05, * * * = p < .001$

## Classification Results

Results of an 'area under the receiver operating characteristic curve' (AUC) analysis of the data used to train the classifiers can be found in Figure 4.4a. Two groups of features corresponding to the MMN and P3a components at fronto-central electrode locations were observed, indicating that these components provide the majority of discriminative information for the binary classification problem.

Mean single-trial classification rates for both groups in all four sessions can be seen in Figure 4.4b. A significant main effect of session on classifier performance [$F(3, 42) = 12.150, p < .001$] was found. Classification rates were significantly higher in session 3 relative to session 1 [$t(15) = -6.228, p < .001$], session 2 [$t(15) = -5.544, p < .001$] and session 4 [$t(15) = 4.910, p < .001$]. Rates were also higher in session 2 relative to session 1 [$t(15) = -2.688, p < .05$]. No effect of participant group was observed, nor was there any interaction effect. Single-trial classification rates thus increased from session to session, with the exception of session 4. To determine whether the drop in performance in

session 4 was due to the restriction of the regularization parameter, another analysis was conducted post-hoc using the full range of regularization parameters. Classification scores in this analysis (Mean = .63) were significantly higher than the rates obtained in session 3 (mean = .62) [$t(15) = -2.929, p = .01$].

The improvement in classifier performance from sessions 1-3 is most likely due to the incremental procedure (i.e. making use of data from all previous sessions) used in training the classifiers. To verify this, the classification analyses were performed for each of the four sessions individually (i.e. non-incrementally). Statistical analysis of these results showed no effect of either group or session, indicating that classifier performance was comparable across both participant groups and across sessions.

## Feedback Values

Mean feedback values from all four sessions with the experimental group are shown in Figure 4.4c. In addition, feedback values were also calculated for the control group during the online sessions, although these values were not in fact used during the experiment. A significant effect of group was found [$F(1, 14) = 5.450, p < .05$]. Mean feedback values were on average higher for the experimental group (.632, St. Err. = .006) than were the simulated values for the control group (.613, St. Err. = .006).

An additional analysis of the simulated feedback values from the control group was performed to determine their relationship to the sham feedback values that were used during the experiment. This was done by correlating the simulated values for all of the standard and deviant trials in which feedback was calculated with the sham values. A very low correlation ($r = .042$) was found. This implies that the sham feedback was uncorrelated with the control participants' ongoing brain responses during the feedback sessions.

## ERPs During Feedback

Averaged ERPs and the corresponding difference waves collected at channels Fz and Cz in the EEG measurements made during the feedback blocks of the four sessions are shown for both groups in Figure 4.5. The corresponding topographies of the MMN and P3a responses can be found in Figures 4.6a and c, respectively. Mean MMN amplitude measurements are plotted in Figure 4.6b. With respect to mean MMN amplitude measurements, a significant main effect of participant group was found [$F(1, 14) = 5.162, p < .05$]. No effect of session or session x participant group interaction was found. It can be concluded that the experimental group showed larger amplitude MMN responses during the feedback blocks relative to the control group.

Mean P3a amplitudes are shown in Figure 4.6d. A main effect of session [$F(3, 42) = 3.934, p < .05$] was found. No effect of participant group or interaction effect of participant group and session was found. A series of pairwise comparisons revealed that, relative to the first session, P3a amplitudes were higher during the third [$t(15) = -2.531, p < .05$] and fourth sessions [$t(15) = -3.136, p < .01$]. The amplitudes in the fourth session were also higher than in the second session [$t(15) = -6.578, p < .001$]. A trend towards larger P3a amplitudes in the experimental group relative to the control group was also observed.

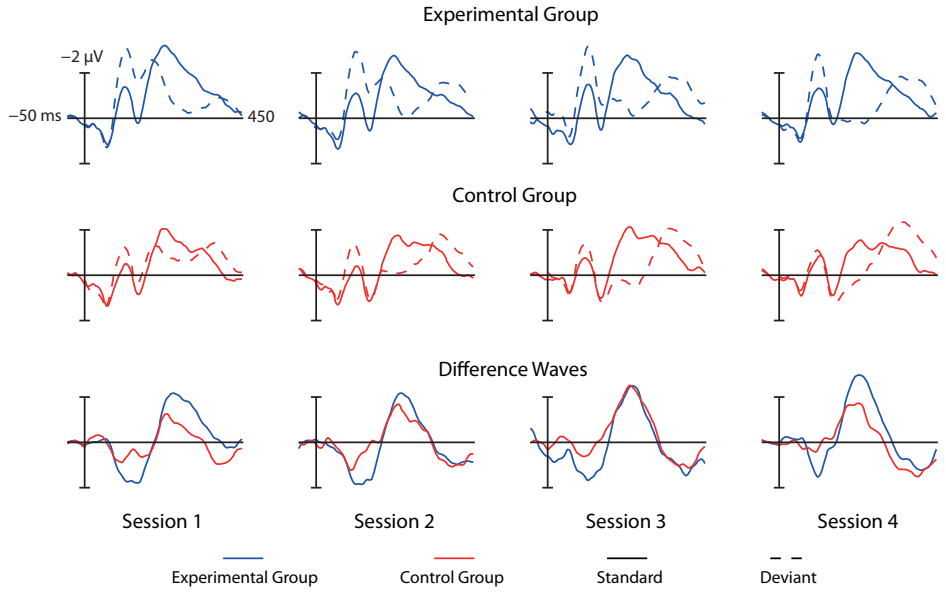


Figure 4.5: *Grand average ERPs and difference waves during feedback blocks.*

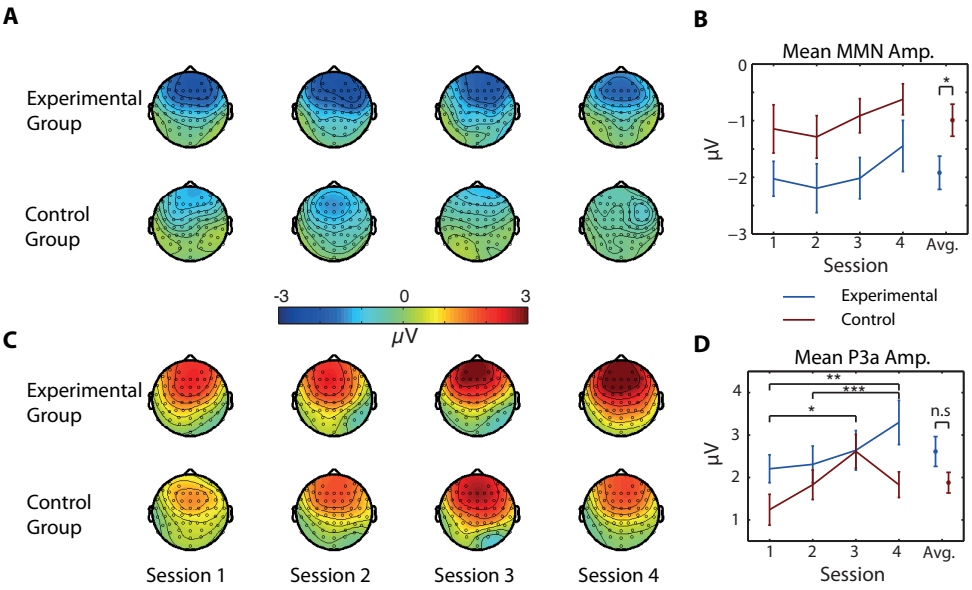


Figure 4.6: *Group-level EEG topoplots and mean amplitude measurements during feedback blocks. A) MMN respsonse topographies. B) Mean MMN amplitudes. C) P3a response topographies. D) Mean P3a amplitudes. n.s. = not significant,* $* = p < .05, ** = p < .01, *** = p < .001$

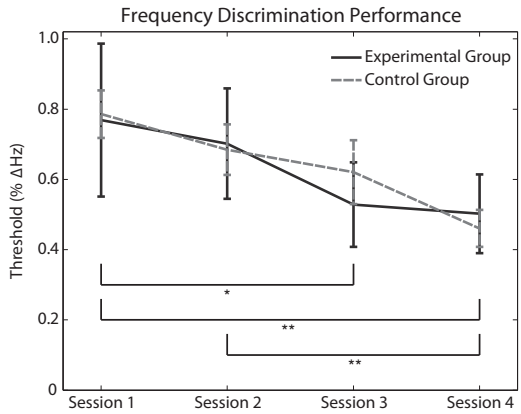


Figure 4.7: *Behavioral discrimination thresholds across sessions. Brackets and asterisks are used to show the results of pair-wise comparisons based on the main effect of session:* $* = p < .05, ** = p < .01$

### Behavioral Discrimination Thresholds

Individual behavioral discrimination thresholds across sessions are shown in Figure 4.7. A significant effect of session was found [$F(3, 42) = 6.642, p < .01$]. The behavioral thresholds were significantly lower following session 4 relative to sessions 1 [$t(15) = 3.431, p < .01$] and 2 [$t(15) = 3.897, p < .01$]. Thresholds measured following session 3 were also significantly lower than in session 1 [$t(15) = 2.185, p < .05$]. No effect of group or group x session interaction was observed. Both groups showed improvements in behavioral discrimination across the four sessions, but there was no difference between the groups.

## Discussion

This experiment was designed to determine if providing participants with neurofeedback based on the decoding of ERPs led to an effect on the ERPs themselves and/or on associated perceptual discrimination abilities. While no difference in the behavior of the two groups was observed, we did see a difference in the brain responses of the two groups during feedback. The current results therefore offer a 'proof-of-concept' for neurofeedback approaches based on pattern classification analysis of evoked responses.

Relative to the amplitude of the initial measurements, both groups showed reduced mean MMN amplitudes during feedback. The brain responses to the sound stimuli had therefore habituated from the first half of each day of the experiment (used to collect classifier training data) to the second half. While such habituation effects are well-known (Ritter et al., 1968; Fruhstorfer, 1971; Sams et al., 1984), they also will play a role in any BCI paradigms requiring the collection of substantial amounts of training data for use in subsequent online sessions. This is because diminished response amplitudes relative to those in the training set data will most likely reduce overall performance (Sellers and Donchin, 2006; Salvaris and Sepulveda, 2009).

However, a change was observed in the two groups from the initial offline measurement

to the feedback measurements. While the MMN response amplitudes measured in the two groups were not significantly different during the initial measurement, this difference was significant during feedback. Additionally, the feedback values presented to experimental group participants were also significantly higher than those simulated for the control participants, even although their overall single-trial performance metrics were not. This suggests that there may have been a positive benefit of feedback on MMN responses due to implicit reinforcement created by the feedback loop. Without necessarily being aware of it, experimental group participants may have adapted the underlying neural responses in such a way that the habituation effect between the first and second halves of each session was less severe than for the control group participants.

This point is supported by a difference in the way participants responded to one of the questions asked following the final session. When asked to quantify the relationship between the sounds and the blurriness, experimental participants had an average response of 2.4 (between 1 'Not at all', and 7 'Very Much', Std. Err. = .71), while control participants all responded with 1. This difference only approached significance ($p = .07$), but nonetheless suggests that experimental group participants perceived a very weak relationship between the basis of the feedback (sound perception) and the feedback itself.

Participants in both groups improved in their behavioral discrimination performance at a rate consistent with the reports of previous longitudinal studies on auditory perceptual learning (Micheyl et al., 2006). The fact that no difference was observed with respect to the neurofeedback suggests that several changes and improvements to the neurofeedback mechanism and procedure are required. First, the stimulus contrast employed was relatively large compared to the individual discrimination thresholds. Altering the stimulus contrast size used during neurofeedback sessions on the basis of individual thresholds could potentially increase the effect of the feedback on perception. Also, the combination of richer stimuli and the use of non-naive, motivated participants could very likely lead to improvements in single-trial classification performance (Höhne et al., 2012; Kleih et al., 2010) and more reliable feedback. The use of cross-participant classification (Herrmann et al., 2012; Brandmeyer et al., 2013) or real-time classifier weight adaptation (Kindermans and Schrauwen, 2013) methods may eliminate the need for the collection of training data, thus bystepping the habituation observed here. Finally, the number of sessions in this study (four) is less than reported in other neurofeedback studies on perceptual learning (Shibata et al., 2011), indicating that more sessions might be needed before an effect would be observed.

The potential for decoded-EEG neurofeedback approaches to enhance perception is constrained by issues such as classifier performance and individual differences that play a role in BCI research as a whole (Wolpaw et al., 2002; van Gerven et al., 2009). Subsequent studies aiming to demonstrate a perceptual learning effect will require further refinement and testing of the neurofeedback protocol. Nevertheless, it can be concluded that these approaches can induce changes in the targeted brain activity, which is a prerequisite for any subsequent perceptual learning applications.

# 5

# Decoding of single-trial auditory mismatch responses for online perceptual monitoring and neurofeedback

*based on Brandmeyer, A., Sadakata, M., Spyrou, L., McQueen, J.M. & Desain, P., 2013. Decoding of single-trial auditory mismatch responses for online perceptual monitoring and neurofeedback. Frontiers in Auditory Cognitive Neuroscience, 7(265), 1-21.*

**Abstract:** Multivariate pattern classification methods are increasingly applied to neuroimaging data in the context of both fundamental research and in brain-computer interfacing approaches. Such methods provide a framework for interpreting measurements made at the single-trial level with respect to a set of two or more distinct mental states. Here, an approach is defined in which the output of a binary classifier trained on data from an auditory mismatch paradigm can be used for online tracking of perception and as a neurofeedback signal. The auditory mismatch paradigm is known to induce distinct perceptual states related to the presentation of high- and low-probability stimuli, which are reflected in event-related potential (ERP) components such as the mismatch negativity (MMN). The first part of this chapter illustrates how pattern classification methods can be applied to data collected in an MMN paradigm, including discussion of the optimization of preprocessing steps, the interpretation of features and how the performance of these methods generalizes across individual participants and measurement sessions. The chapter then goes on to show that the output of these decoding methods can be used in online settings as a continuous index of single-trial brain activation underlying perceptual discrimination. It concludes by discussing several potential domains of application, including neurofeedback, cognitive monitoring and passive brain-computer interfaces.

# Introduction

The ability to non-invasively measure real-time changes in the patterns of brain activity underlying important perceptual and cognitive processes has led to breakthroughs in areas that were until recently the domain of science fiction. These approaches use systems that analyze neuroimaging measurements (e.g. EEG, MEG, fMRI, PET, fNIRS, etc.) as soon as the necessary data is available, such that results can be used online or in real time. As our understanding of the relationship between various forms of brain activity and specific types of mental states and cognitive processes has grown, so too have the number of potential applications of this knowledge in clinical, medical and educational settings (Sellers and Donchin, 2006; Varma et al., 2008; Tzovara et al., 2013; Sellers, 2013). Here, a method for online tracking of brain activity underlying auditory perceptual discrimination is presented. This method is based on the decoding of single-trial auditory evoked potentials, and is illustrated using two datasets collected with variants of the mismatch negativity (MMN) paradigm (Näätänen et al., 2007; Duncan et al., 2009).

In contrast to the averaging methods often used to investigate brain responses measured in EEG, fMRI and other neuroimaging modalities, real-time tracking methods enable researchers to monitor the ongoing dynamics of brain activity as individuals perform different cognitive or behavioral tasks, to use brain responses as a control signal in a brain-computer interface (BCI) setting, or to provide individuals with neurofeedback based on real-time measurements. For instance, real-time fMRI measurements have allowed researchers to develop methods for communication with locked-in patients and patients in vegetative states, as well as novel forms of lie detection and neurofeedback paradigms that help individuals with chronic pain to alleviate some of their symptoms (de Charms, 2008). Others have developed techniques for monitoring working memory function and cognitive load using EEG-based measures (Smith et al., 2001; Brouwer et al., 2012). Single scan dynamic molecular imaging is based on PET measurements and allows for the detection of dopamine release during task performance (Badgaiyan, 2013).

Common to many of these approaches is the use of multivariate pattern classification methods, or so-called decoding approaches (Haynes and Rees, 2006; van Gerven et al., 2009; Blankertz et al., 2011). These machine learning techniques provide a means for making predictions about the mental state of a user on the basis of single-trial neuroimaging data. Predictions are made using a statistical model of a dataset, referred to as a classifier. The dataset used to create the model contains repeated measurements of brain responses corresponding to two or more distinct mental states that are observed in a given task setting. The classifier is trained to identify specific features in the data (e.g. fMRI voxels in regions of interest, EEG samples at specific channels and time-points) that provide discriminative information about the distinct classes of mental states that have been defined as part of the classification problem. Once trained, a classifier can make predictions about novel, previously unseen data.

A clear example of this type of classification problem is provided by BCI systems that use EEG measurements and paradigms designed to elicit the P300 response (for a review of the ERP literature, see Polich, 2007; for P300-based BCIs, see Farwell and Donchin, 1988; Nijboer et al., 2008; Schreuder et al., 2010; Belitski et al., 2011; van der Waal et al., 2012). The P300 response is elicited using sequences containing a rare target event randomly embedded in a series of non-target events. The presentation of a target will draw the user's attention, and is reflected in the P300 component. Data collected in this

paradigm can be thought of as belonging to two classes: targets and non-targets. As the P300 response is only elicited by targets, data collected on target trials will contain the P300 response while non-target trials will not. Given sufficient amounts of data, a classifier trained on such a dataset will learn to assign importance to specific features of the data corresponding to P300 responses elicited in individual trials while ignoring other features unrelated to the two classes of interest. Such a classifier can be used in various BCIs, such as those implementing communication devices or menu systems.

In the context of auditory perception, similar sequences are used to elicit another ERP component: the MMN response. These 'oddball' sequences contain frequent standard trials and rare deviant trials, each corresponding to a different type of sound. For example, the standard sound might be a musical tone with a specific fundamental frequency ($f0$), while the deviant sound has a different $f0$. Whereas the P300 response is elicited using an active task (attend to targets), the MMN is elicited without attending to the stimuli; participants instead watch silent films while oddball sequences of sounds are presented. In the auditory ERP, the MMN is usually maximal at fronto-central electrode locations and peaks between 100-300 ms, depending on the type of stimuli employed (Näätänen et al., 2007). Examples of the ERPs elicited in an MMN paradigm can be found in Figure 5.1.

Interestingly, the MMN response reflects individual differences in perceptual discrimination abilities, such as those related to native-language background (Näätänen et al., 1997; Winkler et al., 1999; see also Chapter Two, published as Brandmeyer et al., 2012) and the effects of musical training (Koelsch et al., 1999; Fujioka et al., 2004), as well as the longitudinal effects of perceptual learning (Tremblay et al., 1997; Menning et al., 2000). Abnormalities in the MMN response are associated with different clinical and medical conditions, such as developmental disorders (Bishop, 2007), schizophrenia (Michie, 2001) and coma (Fischer et al., 2004). This diagnostic aspect of the MMN has led to an interest in the use of decoding methods to analyze single-trial MMN responses. For instance, it has been shown that decoding performance reflects differences in categorical speech perception by native and non-native speakers (see Chapter Three, published as Brandmeyer et al., 2013), and that decoding analyses can be used to predict survival rates in comatose patients (Tzovara et al., 2013).

The ability to track ongoing MMN responses in real-time can provide novel insights into the dynamic nature of perceptual processes. Specifically, real-time monitoring of MMN responses would provide insights into both short- and long-term changes in brain responses associated with perceptual learning. Furthermore, the same real-time tracking technique could serve as the basis of a neurofeedback paradigm centered on auditory perceptual learning, by providing users with ongoing feedback on brain responses associated with discrimination sensitivity. The remainder of this chapter defines a real-time tracking approach based on single-trial decoding of auditory evoked potentials using a logistic regression classification algorithm. First, important aspects of the decoding approach are illustrated, including an outline of the pattern classification problem, data preprocessing, feature evaluation and the generalization of classifier performance. Then a method for interpreting the online classifier output as a continuous signal for use in real-time applications is presented.

Two datasets are used throughout to illustrate various aspects of the decoding and real-time tracking method: one collected using a standard MMN oddball paradigm (Näätänen et al., 2007; Duncan et al., 2009), and one collected using a so-called optimal MMN paradigm (Näätänen et al., 2004). All data were collected using a passive listening task

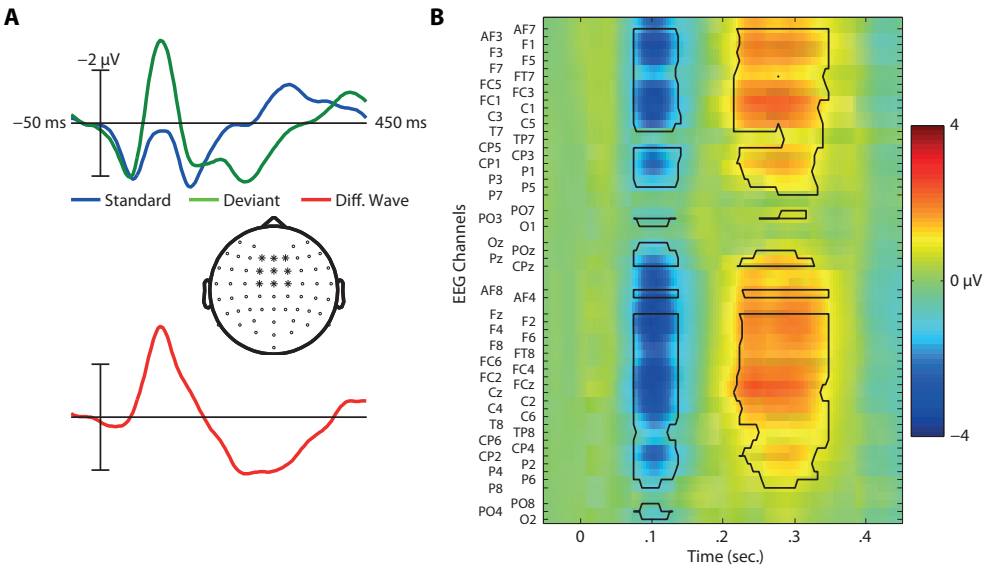


Figure 5.1: *Grand average ERPs and statistical analysis for Dataset 1. A) Grand average ERPs of 14 participants in Dataset 1 for both standard and deviant trials are shown with the deviant minus standard difference wave used to quantify the mismatch negativity response. ERPs were calculated using the average of 9 fronto-central electrode locations (indicated in the scalp map) where the MMN response is typically maximal in amplitude. Relative to the ERP for the standard trials, the deviant ERP shows an enhancement of the negative peak in the N1 time window and an additional positivity between 200-300 ms relative to stimulus onset. This component is referred to as the P3a, and is elicited in MMN paradigms that employ large stimulus contrasts. B) Statistical analysis of the group level responses. Grand average ERPs for all 14 participants were analyzed using a nonparametric cluster randomization procedure (Maris and Oostenveld, 2007) across all 64 EEG channels for time points between 80-350 ms relative to stimulus onset. The results of this analysis are presented using an image of the grand-average difference wave. Two significant clusters of activity (outlined in black) were found, corresponding to the MMN ($p < .001$) and P3a ($p < .001$) responses. The brain responses underlying these components represent two stages in the automatic sensory discrimination process triggered by the presentation of the deviant stimulus.*

in which participants viewed silent films while auditory stimuli were presented. For both datasets, simple tone stimuli were used to elicit the MMN response. Such stimuli are widely used in auditory research, as well as in MMN research. Such tones elicit ERPs containing components whose timing and spatial topography are well understood. A summary of the two datasets can be found in Table 5.1. A complete description of the methods used to collect these datasets is provided in Appendix A.

## Single-trial decoding of auditory ERPs containing the MMN response

The use of pattern classification to decode single-trial EEG data containing different types of task-related brain responses is a hallmark of non-invasive BCI systems (van Gerven et al., 2009) and EEG-based decoding analyses (Schaefer et al., 2011; Brandmeyer et al., 2013). Single-trial analysis and classification of ERP components can be understood in terms of spatial and temporal patterns in the data that are associated with one or more components (Blankertz et al., 2011). Here, the focus lies specifically on a binary classification problem in which the aim is to predict whether a specific single-trial ERP measurement represents a standard (no MMN) or deviant (MMN) trial. The oddball sequence used to collect Dataset 1 is illustrated in Figure 5.2a. Only standard trials which immediately preceded a deviant trial were selected for inclusion in the classification analysis. This balanced the amount of data for each of the two trial types, and is the same as the approach taken in Chapter Three. More recent MMN studies have made use of an 'optimal' MMN paradigm, which employs a sequence structure containing multiple deviant stimuli, and which is illustrated in Figure 5.2b. Dataset 2 was collected using this type of sequence. As an equal number of standard and deviant stimuli are presented, all trials were utilized in the analysis.

The data epoch associated with each trial can be represented as an $m*n$ matrix, where $m$ is the number of EEG channels that are measured (the spatial dimension) and $n$ is the number of recorded samples (the temporal dimension). The data presented here were collected using $m = 64$ EEG channels and $n = 128$ samples, representing data collected between -50 and 450 ms relative to stimulus onset at a sample rate of 256 Hz. Given a labeled dataset (typically '-' and '+') containing examples of standard and deviant trials, a classifier is trained such that each dimension of the corresponding data is assigned a weight, stored in an $m * n$ weighting matrix $w$. The weights are adjusted on the basis of the training data with respect to differences in the observed spatio-temporal patterns associated with the two types of trials, while also taking into account variance in the data not associated with either of the two classes. This latter aspect of the classification problem is crucial in determining the performance of the analysis, given the low signal to noise ratio of MMN measurements (approx. $1 - 5\ \mu V$ for the MMN response compared to ongoing EEG activity, which can exceed $\pm 30\ \mu V$, Duncan et al., 2009; Handy, 2005).

The method used to obtain the weights $w$ determines the type of classifier which is obtained. Methods such as linear discriminant analysis (LDA), support vector machines (SVM) and logistic regression represent just a few of the most common algorithms used. Additionally, the classifier is said to be either linear or nonlinear. The choice to use a linear or nonlinear classifier is often based on the amount and type of data being analyzed, with linear algorithms often preferred for their simplicity (Muller et al., 2003). In the present analysis, which aimed at continuous monitoring of EEG signals, a linear logistic regression

| **Common Recording Methods** | |
|---|---|
| EEG System | Biosemi ActiveTwo amplifier with 64 EEG channels w/ horizontal and vertical EOG, left and right mastoid leads. Data recorded at a sample rate of 2048 Hz and downsampled offline to 256 Hz. |
| Stimulus Presentation | Etymotic ER-4P insert headphones calibrated to approximately 70 $dB$ SPL |
| Data Epochs | Data epoched between -200 to 600 ms (relative to stimulus onset) for preprocessing, data between -50 to 450 ms (non-overlapping epochs) used for ERP and classification analysis. |
| **Dataset 1** | |
| Participants | 14 normal hearing adults |
| Stimuli | Pure sinusoidal tones at 500 Hz (standard) and 600 Hz (deviant), 100 ms duration |
| Sequence Design | Oddball sequences with 85% standard stimuli, 15% deviant stimuli, interstimulus interval of 500 ms, 1000 total trials per block (150 deviants). Based on Duncan et al. (2009). |
| # Blocks per Session | 2 |
| # Sessions | 1 |
| **Dataset 2** | |
| Participants | 12 normal hearing adults |
| Stimuli | Harmonic sinusoidal tones at 500 Hz, 75 ms duration with the following deviant stimuli: Location ($\pm$800 $\mu S$ inter aural delay), Frequency (10% increase in $f0$), Amplitude ($\pm$10 $dB$), Duration (25 ms) and Gap (insertion of 25 ms of silence). |
| Sequence Design | Optimal MMN sequences with 50% standard stimuli and 10% of each of the five types of deviant, alternating standard and deviant stimuli, interstimulus interval of 500 ms. 600 total trials per block (300 deviants). Based on Näätänen et al. (2004), Duncan et al. (2009). |
| # Blocks per Session | 3 |
| # Sessions | 3 |

Table 5.1: *Summary of Datasets 1 and 2.*

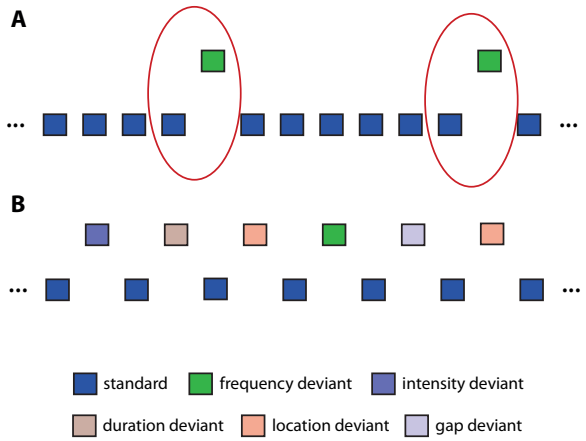


Figure 5.2: *Sequences used during data collection. A) Example of oddball sequences used to collect Dataset 1. An oddball sequence with a 500 Hz standard stimulus (85% of trials) and a 600 Hz deviant stimulus (15% of trials) was used to collect Dataset 1. Only data from deviant trials and the standard trials immediately preceding them were used in the classification analyses (illustrated using red circles). Each block contained a total of 1000 trials. B) Example of 'optimal' MMN sequences used to collect Dataset 2. An alternative sequence structure utilizing 5 different types of deviant stimuli can be used to reduce the overall amount of time required to collect MMN responses (Näätänen et al., 2004; Duncan et al., 2009). Here, all standard and deviant trials are utilized in the classification analysis. Each block contained a total of 600 trials.*

algorithm was chosen for a number of reasons. Firstly, logistic regression classifiers have been shown to provide relatively high performance for EEG data in the context of BCI applications (Farquhar and Hill, 2013). Secondly, the output of logistic regression classifiers can be interpreted probabilistically, as opposed to the output of SVMs and other classifier families (Bishop, 2009). This enables a distinct set of applications which are discussed later.

Given a trained classifier with weights $w$ and an epoch of data $x$, both of which are vectorized versions of the corresponding matrices, the output of a linear classifier is calculated as follows:

$$f(x) = w^\top x + b \tag{5.1}$$

where $b$ is a bias term. This output is referred to as a 'decision value'. For logistic regression classifiers, the range of $f(x)$ is $(-\infty, \infty)$, with the sign of $f(x)$ representing a prediction about which of the two classes in the binary problem $x$ belongs to. A classifier's performance is determined by quantifying the accuracy of predictions it makes about previously unseen data. In other words, we want to know how well a classifier will generalize to novel situations, such as the real-time applications discussed in the introduction. To determine generalization, a given dataset is typically split into two subsets: a training set and a test set. The training set (also referred to as a calibration set) contains data that are used to construct the classifier, while the test set is used to determine how accurately the classifier can make predictions about novel exemplars; because the labels have been provided for epochs in the test set, classifier predictions can be compared with the true

class of the individual exemplars, and the percentage of correct decisions can be calculated. This is referred to as the classification rate.

Rather than estimating generalization using a single training and test set, cross-validation methods can be used. Following the creation of $x$ partitions of the dataset, $x$ iterations of the classification analysis are carried out. In each iteration, a distinct set of $x - 1$ partitions are used as the training set while the remaining partition is used as a test set. Each of the partitions is thus used as a test set in one of the analyses. The classification rates presented here reflect the results of a 10-fold cross-validation procedure, unless otherwise specified. Thus, in each of the 10 iterations, 90% of the available data is used for training, while the remaining 10% is used for testing. The final classification rate obtained in such an analysis, referred to as a cross-validation rate, is the mean of the classification rates obtained for each of the 10 test sets.

An important relationship is defined between the classification rate and the number of classes in a given classification problem. In the case of a binary classifier, imagine that, instead of a classifier, we merely had a random number generator spitting out 1s and 0s (i.e. '+' and '-', the labels of each of the two classes) and making predictions about our data. This is equivalent to the problem of predicting a fair coin toss. In both cases, given infinite trials, we would expect correct answers 50% of the time. This value is defined as chance performance. A binary classifier performing at chance level is essentially no more useful than a random number generator or flipping a coin.

In real-world classification problems, it is rarely possible to achieve 100% classification accuracy. However, poorly defined classification problems can lead to chance-level performance. It is possible to determine whether the results of a given binary classification analysis are significantly different from chance-level performance using binomial confidence intervals (Müller-Putz et al., 2008; Pereira et al., 2009). This is, for instance, the same manner in which the error of a poll is determined, and is based on the number $k$ of observations that are available. A confidence interval $I$ is defined around a specific value $p$ (in the case of a binary problem, chance-level performance, where $p = 0.5$) as follows:

$$I = p \pm z_1 - \frac{1}{2}\alpha\sqrt{\frac{1}{k}p(1-p)} \tag{5.2}$$

where $z_1 - \frac{1}{2}\alpha$ represents a z-scored percentile from a normal distribution for a specific error-value $\alpha$. Importantly, this implies that the confidence interval becomes smaller as $k$ grows. For example, for a statistical confidence level of $\alpha = .05$, the confidence interval for 10 observations is 25.1%, but is only 9.6% for 100 observations and 3.1% for 1000 observations. The observation of classification rates significantly above chance level implies that predictions made by the classifier are non-random. This in turn suggests that the classifier training data contains information that is useful for distinguishing new examples of the different classes from one another, even if this information is not completely reliable.

An alternative to the binomial confidence interval is the permutation test, which can be used in cases where the assumption of class independence does not hold, or when there is reason to suspect bias. Rather then testing the null hypothesis using the binomial distribution, the observed results are compared to the distribution obtained by repeatedly permuting the true class labels belonging to the data and recalculating the classification performance. The probability of the null hypothesis is then calculated as the proportion of the resulting distribution with classification performance greater than or equal to the observed result Pereira et al. (2009). In general, however, the significance of classification

results presented here are based on the use of binomial confidence intervals, assuming chance-level performance of 0.5.

A potential problem arising in classification analyses is the over-fitting of the classifier to the training data. What this means is that the model does not generalize well to new examples outside the training set. This is especially troublesome for high-dimensional data sets such as those encountered in neuroimaging. For instance, the present data contains 64 x 128 = 8,192 dimensions. A rule-of-thumb in pattern classification is that the number of examples needed to train a classifier is roughly equal to the number of dimensions in the data (Duda et al., 2001; Blankertz et al., 2010). However, this is very often impractical in the case of neuroimaging data due to the amount of time required to obtain suitable training data.

One solution for dealing with over-fitting is the use of regularization methods. These methods limit the complexity of classifier models, as over-fitting is associated with relatively more complex models. For logistic regression classifiers, the level of regularization is a function of the total variance observed in the training data. A weighting parameter $c$ determines classifier complexity, with smaller values leading to more complex models. The optimal value of $c$ is determined using a grid search with the values $[.001, .01, .1, 0, 1, 10, 1000]$, respective to the overall variance of the data: the 10-fold cross-validation procedure is repeated for each value of $c$, and the model trained with the regularization setting leading to the highest overall performance is selected.

A principal requirement of classifiers intended for use in real-world applications is that their performance should be as high as possible. Prior to classifier training, pre-processing and/or feature selection steps are typically performed. These steps aim to remove noise from the data and to reduce its dimensionality. The precise nature of the steps taking during pre-processing and feature selection is determined by the type of data. Additionally, the amount and type of data available for classifier training can significantly impact their performance.

To summarize, the present approach uses pattern classification methods to make predictions about single-trial auditory ERPs collected in an auditory mismatch paradigm. Specifically, the analysis makes use of 500 ms epochs of 64-channel EEG data from two types of trials, standard and deviant, to define a binary classification problem. A regularized, linear logistic regression classifier is then trained using equal amounts of data from both types of trials. A 10-fold cross-validation is used initially to estimate generalization performance. The following sections present a data-driven approach in which preprocessing parameters for classification of MMN responses are optimized, and examine the effects of training set size and data selection on classifier performance. Relative changes in classifier performance between analysis steps are evaluated using paired-samples t-tests. Methods for evaluating data features that influence classification performance are also presented, followed by an assessment of generalization performance using datasets collected from different measurement sessions or individuals.

## Pre-processing

In the case of EEG data, pre-processing is also a typical step in ERP analyses, and many of the specific actions taken, such as artifact removal, filtering and resampling, remain the same. However, due to the nature of the pattern classification problem, the parameters used during these steps may differ from those employed in ERP paradigms. For instance,

the cutoff frequencies and downsampling employed by decoding approaches are often more severe due to the influence of noise and brain activity unrelated to the primary task on classifier performance. Recent work has evaluated pre-processing methods for obtaining optimal classification performance with EEG data collected in different BCI paradigms (Farquhar and Hill, 2013). Here, a similar evaluation is performed, the results of which show that the optimal parameters are somewhat different for data collected using an MMN paradigm.

The datasets presented here were initially preprocessed using the following steps: 1) Bad channels (i.e. those with offsets greater than 35 $mV$ or with 50 Hz power greater than 1000 $\mu V^2$) were repaired using a spherical spline interpolation procedure (Perrin et al., 1989) on a per epoch basis. This has the effect of removing artifactual and noise-related activity that can influence classification performance, and preserves the full electrode montage for subsequent steps in the decoding analysis. 2) An independent component analysis was performed, and components containing artifactual activity were selected using a threshold calculated on the basis of the mean variance across trials and components. These selected components were removed from the data, and the remaining components were re-projected onto the original EEG recording channels. The same method has been employed in Bishop and Hardiman (2010) and Chapter Three in the context of MMN paradigms, and was originally described in Jung et al. (2000). This has the effect of removing eye movements and other muscular artifacts that would otherwise lead to the rejection of data epochs in subsequent steps. The preservation of individual epochs is important for optimizing performance, as increased training set size leads to higher classifications rates. 3) Data were band-pass filtered between 1-25 Hz (Kujala et al., 2007). 4) Individual epochs were inspected at all EEG channels for activity exceeding $\pm 75\ \mu V$, in which case they were excluded from subsequent analyses. 5) Lastly, data were re-referenced to the average of the two mastoid leads, which is known to enhance the signal-to-noise ratio of MMN responses (Kujala et al., 2007; Duncan et al., 2009).

Figure 5.3a presents average cross-validation rates obtained with Dataset 1 when training classifiers on individual datasets following each preprocessing step. Significant improvements in performance were obtained when the data were filtered [$t(13) = -3.026, p = .01$]. Are additional improvements possible through more refined selection of cutoff frequencies? Figure 5.3b presents an assessment of the effects of high- and low-cutoff frequencies. Maximal performance was obtained with high- and low-pass frequencies of 0.5 and 13 Hz, respectively. A comparison of performance using different filters is presented in Figure 5.3c. Performance was slightly higher relative to the original 1-25 Hz band-pass filter when using a 0.5-13 Hz band-pass filter, as well as relative to a 1-20 Hz band-pass filter as recommended for MMN recordings in Duncan et al. (2009). Research investigating the oscillatory activity underlying the time-domain MMN response has revealed that phase resetting and power modulation of theta band activity (4-8 Hz) in the temporal and frontal MMN generators, respectively, occur during deviant trials (Fuentemilla et al., 2008). This suggests that activity in and around this frequency band will provide discriminative information for the present classification analysis. Other studies classifying MMN data (Herrmann et al., 2012) and auditory evoked responses to music (Schaefer et al., 2011) have also made use of similar filter settings during preprocessing, indicating that band-pass filtering in the 0.5-13 Hz range might improve decoding performance for auditory brain responses.

The choice of reference electrodes represents another preprocessing step where multiple choices exist that might influence overall classification performance. This was evaluated by

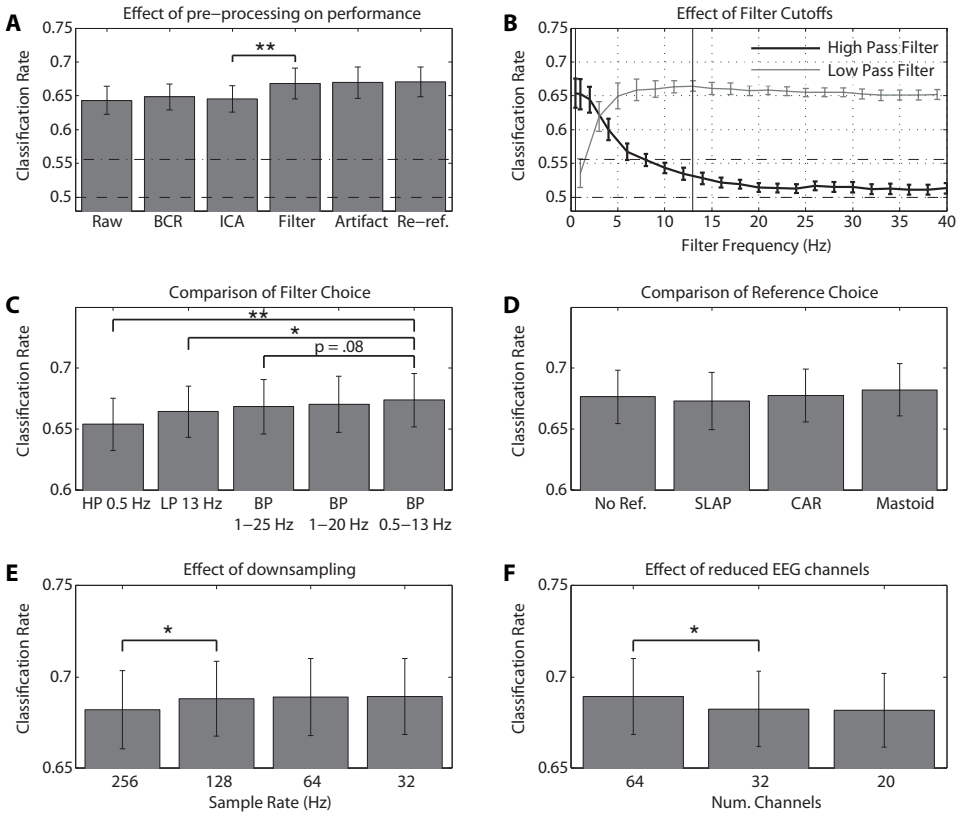


Figure 5.3: *Effects of preprocessing on cross-validation performance for Dataset 1. A) Comparison of average cross-validation rates across subsequent stages of preprocessing. Individual participant's data were used to train a linear logistic regression classifier following each of five preprocessing steps: bad channel repair (BCR), independent component analysis based artifact removal (ICA, Bishop and Hardiman, 2010), band-pass filtering between 1-25 Hz (as suggested in Kujala et al., 2007), rejection of data epochs containing activity exceeding* $\pm 75\ \mu V$*, and re-referencing the data to the average of the two mastoid leads (Kujala et al., 2007; Duncan et al., 2009). Baseline correction between -50 to 0 ms (relative to stimulus onset) was applied at each step. As is illustrated, a significant increase in performance was obtained during the filtering stage. B) Evaluation of filter cutoff frequencies. Two separate grid search procedures were used to evaluate the effect of high- and low-pass cutoff frequencies on performance. Optimal cutoffs of 0.5 Hz and 13 Hz were found for the low- and high-pass filters, respectively, and are illustrated in the plot. C) Comparison of filters. Classification rates using the optimal low- and high-pass filter parameters as well as the original 1-25 Hz band-pass filter (Kujala et al., 2007), a 1-20 Hz band-pass filter as recommended by (Duncan et al., 2009) and a 0.5 - 13 Hz band-pass filter based on the previous analysis are shown. Performance was highest using the 0.5 - 13 Hz band-pass filter. D) Comparison of reference choice. Cross-validation rates are shown for unreferenced data as well as for three possible reference choices: common-average reference (CAR), a surface laplacian reference (Farquhar and Hill, 2013) and a mastoid reference.* *Continues on next page...*

Figure 5.3: *continued... Although no significant difference in classifier performance was found, performance was highest overall when using a mastoid reference. As has been previously discussed in the literature, the use of a mastoid reference can enhance the signal-to-noise ratio of auditory ERPs collected using an MMN paradigm (Kujala et al., 2007; Duncan et al., 2009). E) Effect of downsampling on cross-validation rates. Classifier performance using the original 256 Hz sampling rate and the updated filter cutoffs was compared to performance using data resampled at 128, 64 and 32 Hz. A significant improvement in average performance was observed for the initial step, with the highest performance observed for data downsampled to 32 Hz. As the high-pass cutoff of 13 Hz is below the nyquist frequency (approx. 16 Hz), these improvements suggest that the removal of unnecessary dimensions in the data has a beneficial effect on classifier performance. F) Effect of reduced EEG channels. A comparison of the initial 64-channel montage with 32- and 20-channel montages based on the international 10-20 system showed a significant decrease in performance. This implies that features of the data related to the topography of the EEG signal have an important bearing on classification performance. For all subfigures, the significance of paired-sample t-test comparisons is indicated using asterisks:* $* = p < .05$, $** = p < .01$

comparing performance when using three different reference montages (surface laplacian reference, common average reference, and averaged mastoid reference) as well as when the referencing step was omitted. The results are presented in Figure 5.3d. No significant differences in overall performance were found, but classification performance was highest overall for the mastoid reference, suggesting that the signal-to-noise benefits provided by a mastoid reference that have previously been described (Kujala et al., 2007; Duncan et al., 2009) might also contribute to higher classification rates.

Two final preprocessing steps were evaluated: additional downsampling of the data, and reduction of the number of electrodes included in the final montage. These steps have important consequences for the classification problem: the reduction of the number of temporal (i.e. number of samples) and spatial (i.e. number of EEG channels) dimensions leads to overall reduction in the dimensionality of the data. As previously mentioned, dimensionality reduction implies that fewer training examples are needed to obtain optimal performance. Another benefit of removing EEG channels is the use of less electrodes during measurements, which saves time during cap-fitting.

The effects of temporal and spatial downsampling on cross-validation rates are presented in Figures 5.3e and f, respectively. A contrasting picture emerges. While performance significantly improves when downsampling from 256 Hz to 128 Hz [$t(13) = -2.360, p < .05$], performance is reduced overall when the number of electrodes is decreased from 64 to 32 [$t(13) = 2.956, p < .05$]. There are several factors that might underly this difference in spatial vs. temporal features. Firstly, in the temporal dimension, the sampling rate determines the Nyquist frequency, and thus the spectral range of activities that are captured in the data. The data has been filtered between 0.5 and 13 Hz, meaning that sample rates of 26 Hz and higher should capture the range of activity in the data. By reducing the sample rate to 32 Hz, the dimensionality of the data has been reduced by 77.5%, from 8,192 dimensions to 1,024. This may impact overall performance. Secondly, the ability of the classifier to optimally represent the spatial structure of both class-relevant signals and class-irrelevant noise might also depend on the relative density of electrode placement. Thus, reduction according to the 10-20 system is potentially suboptimal. Several previous studies have investigated the use of spatial filtering techniques (i.e. weighting of specific electrode locations) in the context of EEG pattern

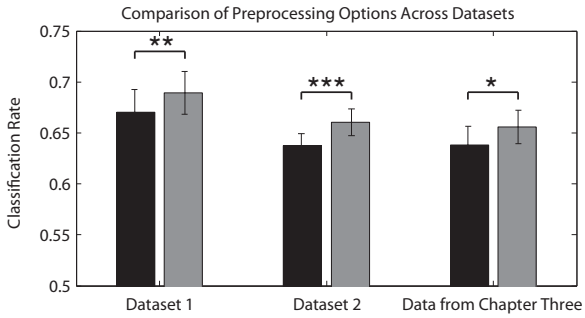


Figure 5.4: *Comparison of preprocessing effects on classification performance across datasets. The two datasets presented in this chapter, as well as data collected in Chapter Three from 11 native English speakers using English language speech stimuli, were used to evaluate whether the optimal parameters for preprocessing steps observed for Dataset 1 generalized to novel datasets and types of stimuli. Performance was compared when using the original filter (1-25 Hz band-pass) and sampling rates (256 Hz for Datasets 1 & 2, 128 Hz for Chapter Three), and when using the optimal filter (0.5 - 13 Hz) and sample rate (32 Hz). Significant improvements in mean performance were observed for all three datasets. The significance of paired-sample t-test comparisons is indicated using asterisks:* $* = p < .05$, $** = p < .01$, $*** = p < .001$

classification. Brunner and colleagues showed that the output of an infomax-based ICA could be used as a spatial filter to improve classifier performance in a dataset collected in an imagined movement paradigm (Brunner et al., 2007). Farquhar and Hill also showed that a technique known as spatial whitening could be used to attain similar improvements in performance as with ICA (Farquhar and Hill, 2013).

Compared to the results obtained with the original preprocessing parameters, classifier performance is significantly improved using the updated parameters $[t(13) = -4.216, p = .001]$. It's worth knowing whether the improvements obtained through changes in filter settings and by downsampling generalize to other datasets collected in an MMN paradigm and using different types of stimuli. This was evaluated with Dataset 2, as well as using a subset of the data presented in Chapter Three, collected in an MMN paradigm from 11 native-English speakers using English phonemes. The analysis in Chapter Three made use of similar initial preprocessing steps and a decoding analysis. Significant improvements in average classification rates were obtained using the current parameters for both Dataset 2 $[t(11) = -4.484, p < .001]$ and for the data from Chapter Three $[t(10) = -2.479, p < .05]$ (see Figure 5.4). This implies that these parameters may serve as useful guidelines for EEG based-decoding approaches using an MMN paradigm. A summary of the optimized parameters can be found in Table 5.2. The remainder of this chapter makes use of data pre-processed using these parameters.

## Effects of collection methods and dataset size

Data collection methods and the amount of available data can also influence classifier performance. For example, ERP component amplitudes will often decrease during measurements due to habituation (Ritter et al., 1968; Fruhstorfer, 1971; Sams et al., 1984). Previous BCI research has demonstrated that such habituation effects can lead to reduced

| Step | Parameters |
|---|---|
| Bad-channel detection and repair | Individual electrodes with offsets greater than $\pm 35\ mV$ or 50 Hz power greater than 1000 $\mu V^2$ repaired using spherical-spline interpolation of neighboring electrodes on a per-epoch basis |
| ICA-based artifact removal | Infomax-based ICA procedure used to obtain independent component transform of EEG data. Individual components with variance above overall mean variance of the dataset selected for removal, visual inspection of selected components followed by reproduction of data to EEG channels |
| Filter | Band-pass between 0.5 and 13 Hz |
| Artifact Rejection | $\pm 75\ \mu V$ |
| Rereferencing | Average of left and right mastoid leads |
| Resampling | 32 Hz |

Table 5.2: *Summary of optimized preprocessing parameters.*

performance (Sellers and Donchin, 2006; Salvaris and Sepulveda, 2009).

Figure 5.5 illustrates this phenomenon using Dataset 2, for which three consecutive measurement blocks were available for each of three separate measurement sessions on different days. On the one hand, looking at average classifier performance across the three blocks of the first session, classification rates progressively decrease, with the difference in performance between the first and third blocks reaching significance $[t(11) = 2.979, p < .05]$. On the other hand, when comparing the initial blocks recorded in each of the three sessions, no significant differences in the average performance rates are observed. This suggests that class-relevant differences in brain responses to standard and deviant trials are relatively enhanced in the initial portions of a given measurement session. In support of this, an analysis of the mean MMN component amplitudes across the three blocks measured in the first session showed a gradual reduction in negativity (First block: -3.90 $\mu V$, second block: -3.71 $\mu V$, third block -3.43 $\mu V$), although these differences were not significant.

Another factor that influences performance is the amount of MMN data used for classifier training. This is an issue shared with ERP paradigms that analyze individual MMN responses. It has been suggested that 200-300 deviant trials should be collected to reliably estimate individual MMN amplitudes, latencies and statistical significance (Bukard et al., 2007; Duncan et al., 2009; Bishop and Hardiman, 2010). Previous analyses looking at the effects of EEG dataset size on classification performance have also shown consistent improvement in classification accuracy as more data becomes available (Blankertz et al., 2011; Farquhar and Hill, 2013).

The effects of incrementally increasing dataset size on cross-validation performance are illustrated in Figure 5.6. For both datasets, performance improves as more data are included. Gains in performance are greatest during the initial increases in dataset size. For Dataset 1, performance significantly improved when the amount of total trials was increased from 100 to 200 $[t(13) = -2.515, p < .05]$. For Dataset 2, significant increases were observed when increasing from 400 to 500 $[t(11) = -2.337, p < .05]$, and from 500 to 600 trials $[t(11) = -3.448, p < .01]$. As Dataset 2 contained a much larger number of total data epochs (up to 5400 per individual participant), it was possible

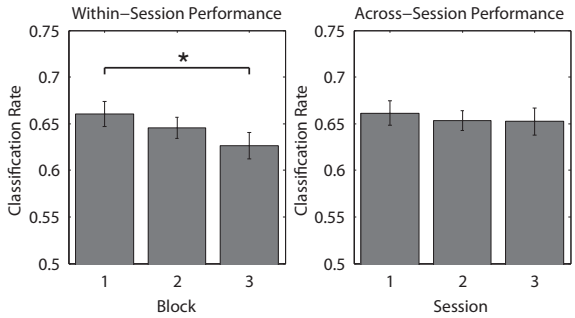


Figure 5.5: *Comparison of performance across measurement blocks and sessions. Classification rates obtained using individual measurement blocks from Dataset 2 were compared. While no significant difference in performance was found for the initial blocks in each of the three sessions, a significant decrease in performance from the first to the third block of the first session was observed. This may be due to a number of factors, including response habituation and experimental fatigue, both of which have been shown to influence ERP classification performance (Sellers and Donchin, 2006; Salvaris and Sepulveda, 2009). Such effects should be carefully considered during the design of experiments and procedures that make use of ERP classification. The significance of paired-sample t-test comparisons is indicated using asterisks:* $* = p < .05$

to examine whether these gains continue when including data recorded across multiple sessions. Additional performance benefits were observed when increasing the amount of data utilized from 600 to 2400 trials [$t(11) = -3.176, p < .01$], and from 2400 to 3600 trials [$t(11) = -6.675, p < .001$]. Indeed, performance was maximal (70% on average) when utilizing all data from the first two sessions or from all three sessions combined. It thus appears that decoding analyses of MMN data from multiple measurement sessions can benefit from pooling data within and across sessions. This benefit seems to outweigh the decrease in performance observed across consecutive blocks measured within a session.

## Evaluation of class-relevant features of MMN data

This section illustrates methods for evaluating specific class-relevant features of these data. In general, when applying classification methods to time-domain EEG data, it is useful to understand which specific ERP components contribute to classifier performance (Blankertz et al., 2011). While one would obviously expect the MMN component to contribute to classifier performance in these analyses, it is not the only component of the auditory ERP modulated during the presentation of a deviant trial. For instance, Figure 5.1 illustrates that the P3a response is clearly elicited by deviant trials. Two different methods are used to evaluate the class-relevant features present in Dataset 1: so-called 'searchlight' methods (Haynes and Rees, 2006; Blankertz et al., 2011; Chen et al., 2011; Herrmann et al., 2012) in the temporal and spatial domains, and area-under-the receiver operating characteristic, or area-under-the-curve (AUC) scores (Fawcett, 2006).

Individual components such as the MMN are defined by both the time intervals in which they occur as well as by the spatial distribution of the corresponding scalp potentials. One method for evaluating the contribution of individual time-points and EEG channels is to

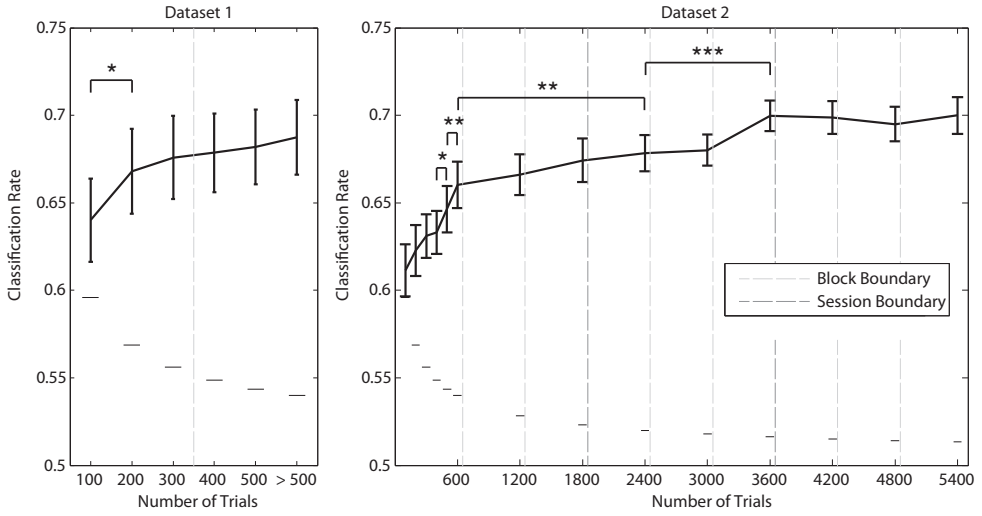


Figure 5.6: *Effects of dataset size on average cross-validation rates. Individual data from Dataset 1 and Dataset 2 were used in incremental steps to train classifiers on increasingly larger amounts of data. The order in which the data was collected was preserved. Block/session boundaries are indicated along with binomial confidence intervals for the maximal number of available data points in each iteration. The amount actually used for some participants was slightly reduced at certain intervals due to the removal of data epochs during the artifact rejection preprocessing step. Data from additional blocks collected for some participants in Dataset 1 were utilized in the final iteration. With respect to Dataset 1, a significant increase in performance was observed for the initial increase in dataset size, from 100 trials to 200 trials. Performance was highest when using the complete dataset. For Dataset 2, data from each of the three blocks in each of the three measurement sessions were incrementally added to the analysis. Significant improvements were observed up to the inclusion of 3600 data epochs (i.e. the full data sets from the first and second sessions). From here on, average classification performance remained at approximately 70%. Error bars indicate the standard error across participants, while the small horizontal lines indicate the binomial confidence interval for the maximum number of data epochs that were available. Consecutive pairs of incrementation steps for which significant improvements in performance were observed are indicated with brackets and asterisks:* $* = p < .05$, $** = p < .01$, $*** = p < .001$

perform a searchlight analysis in the temporal and spatial domains. When a searchlight analysis is performed in the spatial domain, a series of classification analyses are performed at each individual EEG channel using all available time-points. Channels at which the various ERP components distinguishing the two types of trials are maximal will show the highest cross-validation rates. Conversely, in the temporal domain, a series of classification analyses are performed on each available time-point using all available EEG channels. Time-points with large differences in the amplitudes of the scalp potentials across channels for the two types of trials will generally show the best performance.

The results of both types of searchlight analyses are presented for two individual participants as well as averaged across all 14 participants from Dataset 1 in Figure 5.7. Figure 5.7a presents results in the spatial domain. Here, the fronto-central channels typically associated with MMN peak amplitude clearly show the highest classification performance, indicating that these channels provide the most information regarding the two classes of ERPs. Conversely, occipital channels show relatively poor performance, which is not surprising, given what is known about the scalp distributions associated with auditory evoked potentials (Bukard et al., 2007).

Figure 5.7b plots the equivalent results for the searchlight analysis performed in the temporal domain. Here, clear peaks in both the individual and average data can be seen for time-points corresponding to the MMN and P3a intervals, and, to a lesser-extent, at time-intervals immediately preceding and following stimulus onset. Interestingly, classifier performance of individual time-points is maximal within the time-interval associated with the P3a interval, which suggests that the attentional shift associated with the P3a component corresponds to single-trial brain responses that are relatively robust compared to the MMN component.

In the section on preprocessing, it was shown that temporal downsampling led to an overall improvement in performance while spatial downsampling had a negative impact on performance. The results of both searchlight procedures can also be used to investigate the effects of gradually increasing the number of channels and time-points utilized in the classification analysis. By sorting these channels and time-points on the basis of both individual and averaged results, an incremental procedure can be used to estimate which subsets lead to optimal classifier performance. This incremental procedure is preferred to a more comprehensive evaluation of electrode combinations due to the very large (approximately $10^{89}$) number of possible permutations.

These results are presented in Figure 5.7c. Analyses based on both individual and average searchlight results converge at relatively similar points: between 7 to 13 channels and 1-3 time-points can be removed without reducing overall performance. Results based on the average searchlight results across participants lead to a slightly smaller number of channels and time-points than the individual results. These channels and time-points are illustrated in Figure 5.7d. In particular, the results of channel removal could potentially be used to reduce the number of electrodes in the optimal cap-montage to 51, a reduction of approximately 20% that could positively impact cap-fitting times in online settings. Further reductions appear to negatively impact performance. It has been suggested that the use of a larger number of EEG channels generally improves the performance of ERP component classification, provided that appropriate regularization methods are used (Blankertz et al., 2010).

Methods also exist for evaluating individual spatio-temporal features of the data. An AUC analysis represents one such approach that has proven useful in the context of clas-

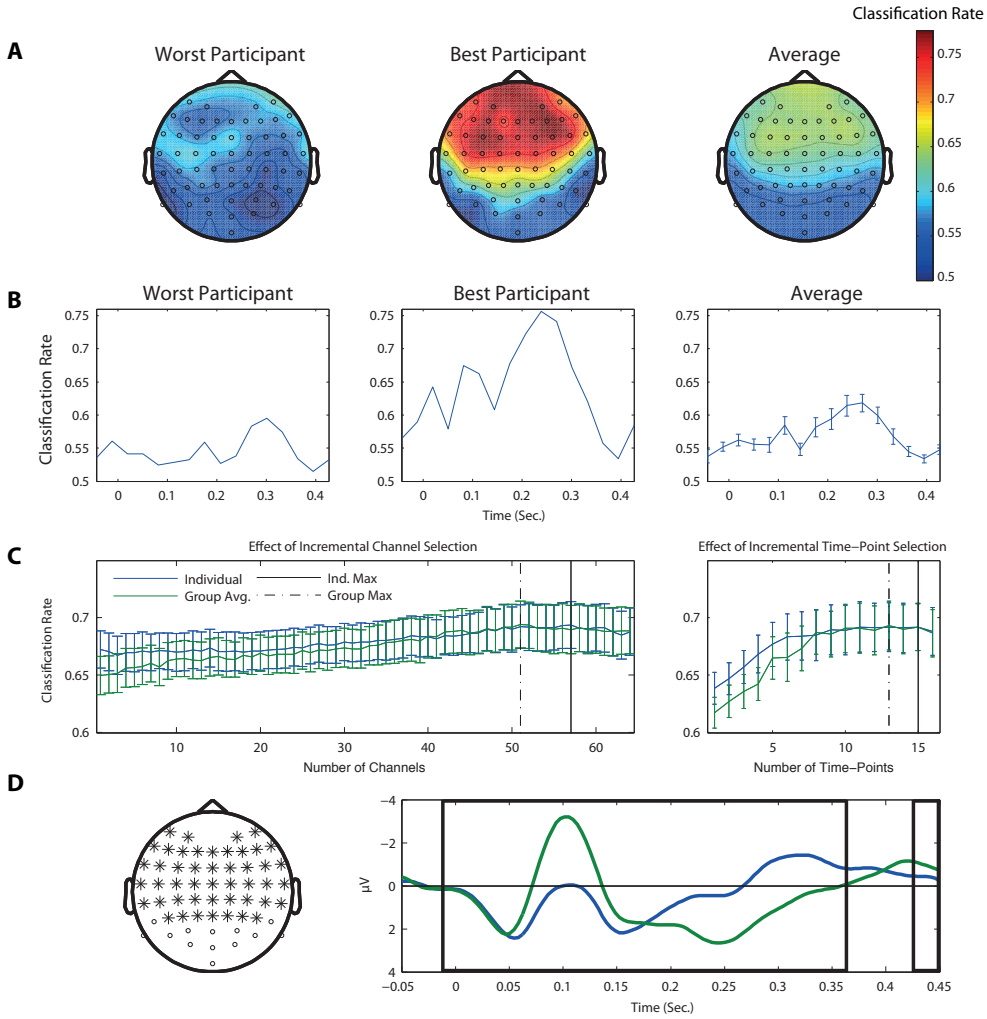


Figure 5.7: *Use of searchlight procedures for spatial and temporal classification of Dataset 1. A) Classifier performance using individual EEG channels. Cross-validation rates at individual EEG channels are shown for the two participants with the lowest and highest cross-validation rates, as well as the average rates across all participants. The fronto-central electrode locations typically associated with maximal MMN responses are where the highest performance was observed. B) Classification performance using individual time-points. Performance for the same two participants and the average across participants are shown. Peaks in performance can be seen for both the MMN and P3a component time-ranges, as well as an additional peak near stimulus onset. When comparing the timing of the peaks for the two individual participants, a relative shift in their latencies can be observed. Similar shifts in ERP component latencies are typically associated with individual differences in perceptual discrimination and task performance. C) Effects of incremental inclusion of channels and time-points based on searchlight performance results. Cross-validation performance rates were evaluated using an incremental procedure in which either individual EEG channels or time-points were included on the basis of their performance ranking in the previous searchlight analyses. This was done using the rankings obtained for both individual participants as well as the average ranking across participants. Continues on next page...*

Figure 5.7: *continued. The peak performance is indicated for both individual and group rankings. D) Selection of optimal channels and time-points on the basis of average rankings. 51 EEG-channels and 13 time-points are highlighted on the basis of the maximal average cross-validation rates obtained using the incremental procedure. The selection of optimal EEG channels can potentially reduce the size of the cap-montage used during measurements, thus reducing the time needed for cap-fitting by approximately 20%, as well as further reducing the dimensionality of the data used during classifier training (from 64 channels x 16 time points = 1024 dimensions to 663).*

sification analysis (Fawcett, 2006). AUC scores are derived from the receiver operating characteristic (also referred to as the ROC curve), which is an analytical tool developed in the context of signal detection theory, and which is also widely used in psychophysics (Green and Swets, 1974). Essentially, the AUC score (between 0 and 1) quantifies the ratio of true-positives to false-positives in a signal detection task. This has proven useful in the context of machine learning and pattern classification, as it provides a richer measure of classifier performance than accuracy alone.

An AUC score of .5 indicates chance level performance, and suggests no discriminative information is available. In the context of the present binary classification problem, we would like to know if a given spatio-temporal dimension of the data provides information about the two classes (i.e. standard and deviant trials). Following an analysis of each individual dimension, the scores can be visualized as a two-dimensional image (i.e. space x time), similar to that used for visualizing the grand-average difference wave in Figure 5.1. Examples of these plots can be found in 5.8a for the same two individual participants as in Figure 5.7, as well as for the entire dataset. Features corresponding to the MMN and P3a components can be seen at the expected channels and time-points.

Similarly to the incremental procedure used to select channels and time-points on the basis of the searchlight analysis, the AUC scores for individual spatio-temporal dimensions can be used to incrementally reduce the overall number of dimensions in the data used to train a classifier. This procedure is illustrated using both individual and group data in Figure 5.8b. Corresponding effects on classifier performance and the data dimensionality are presented in Figure 5.8c. Similarly to the searchlight procedure, the results of this analysis suggest that specific channels and time-points can be removed from the data without affecting overall performance.

While searchlight and AUC methods both offer a means of evaluating features that contribute to classifier performance, their use in selecting features for removal does not lead to any substantial gains in performance relative to the performance obtained following the optimization of preprocessing parameters. This is due to the fact that both methods are essentially univariate in nature, while the classifier training methods are multivariate. In essence, they fail to capture specific patterns in the covariance of multiple features useful for distinguishing class-relevant signals mixed with class-irrelevant noise. In contrast, the gradient descent and regularization methods employed during classifier training to determine the optimal weighting of individual features have in fact already selected the most informative features of the data. However, both methods provide tools for understanding which features of the data contribute to classification performance.

One final method of selecting features for classification is based on the cognitive mechanisms underlying the generation of specific ERP components. In the case of auditory

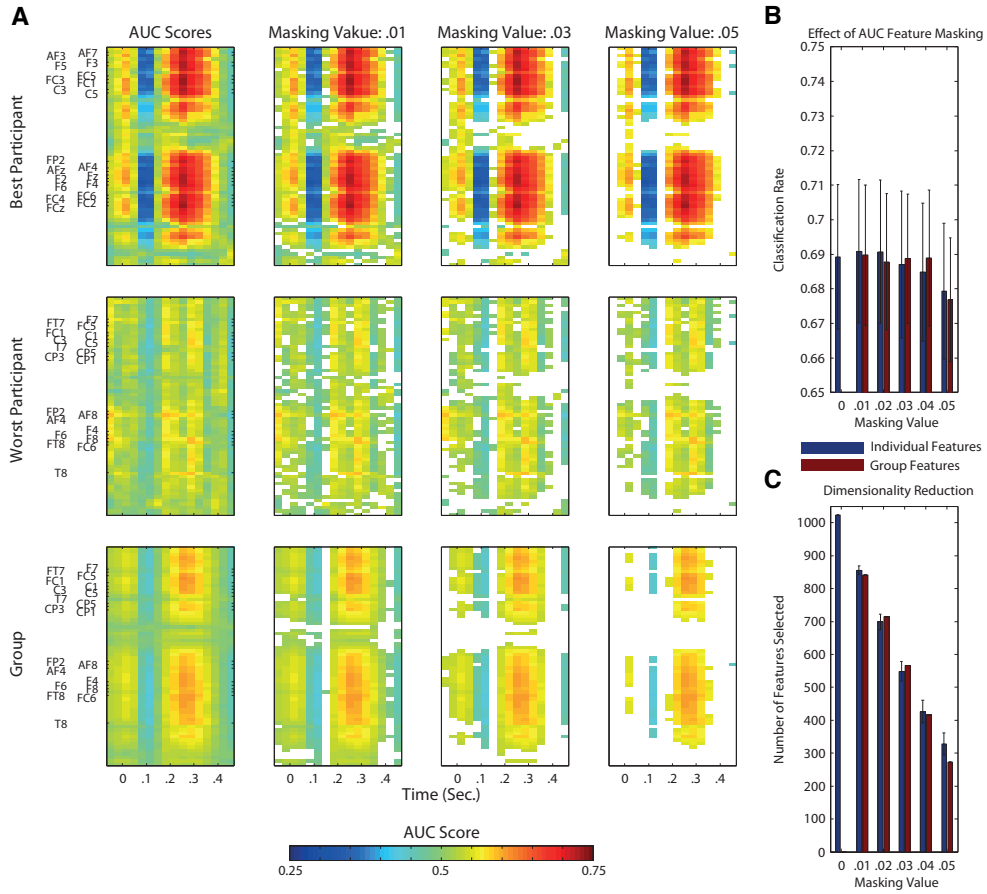


Figure 5.8: *Area under the receiver operating curve (AUC) scores and masking procedure. AUC scores are often used to evaluate the class-relevant information present in datasets used for classifier training. Additionally, they can be used to select the most relevant dimensions of the data during classification analysis. A) Such a procedure is illustrated here using the same two individual participants as in Figure 5.7 as well as using all participant data from Dataset 1. Individual dimensions were selectively masked (i.e. removed from the dataset) on the basis of their AUC score. This masking value was defined as* $|AUC - .5|$*, that is, the absolute value of the AUC Score minus .5. This is effectively a measure of a feature's distance from the no discrimination line of the ROC plot. Additionally, this analysis was performed on the basis of AUC scores calculated using both the individual and the group data. B & C) Using features selected on the basis of the group data and a masking value of* $.04$*, it is possible to reduce the dimensionality of the data by over half (from 1024 to approximately 400, shown in panel C) without a reduction in average performance (see panel B). The remaining dimensions correspond approximately to the two clusters observed in the statistical analysis presented in Figure 1 (i.e. MMN and P3a components) as well as to additional activity around stimulus onset and between 400-450 ms following stimulus onset.*

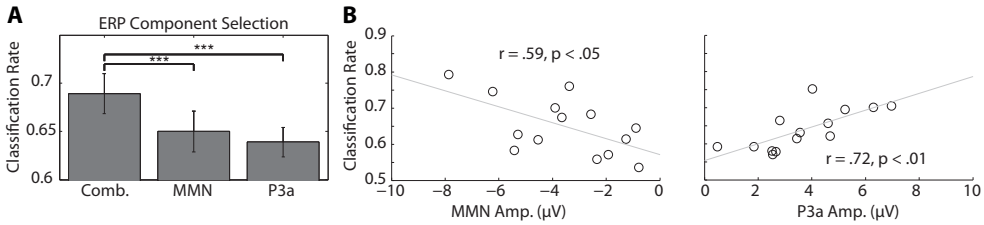


Figure 5.9: *Classification of individual ERP components. A) Two classification analyses were performed using individual participant's data from Dataset 1. Subsets of the data from the time windows associated with the MMN (82-207 ms) and P3a (238-363 ms) were selected and classified. Relative to analyses using the entire time range (-50-450 ms), cross-validation rates in both of these analyses were significantly lower, indicated using asterisks:* $***= p < .001$ *B) Correlation analyses of the individual results with the ERP component amplitudes measured using the individual difference waves were significant for both the MMN and P3a, indicating that classifier performance is strongly related to the amplitude of these components.*

mismatch data, two principal components distinguish the standard and deviant trials: the MMN and the P3a. While the MMN reflects pre-attentive sensory processes that occur in bilateral auditory cortex, the P3a component reflects subsequent activity generated in frontal cortex that is associated with an attentional switching mechanism (Näätänen et al., 2007; Polich, 2007). While these processes are essentially intertwined, they reflect two distinct forms of brain activity related to perception and attention. In the context of different perceptual tracking and neurofeedback applications, it may be the case that the decoding analysis should focus on one or the other component, depending on which specific cognitive processes are deemed to be most relevant.

Figure 5.9a presents the results of two additional classification analyses performed using only data in the time intervals associated with the MMN and P3a responses, relative to the performance obtained using the entire 500 ms window (-50-450 ms) previously established. Classifier performance significantly decreased for both the MMN $[t(13) = 4.715, p < .001]$ and the P3a $[t(13) = 4.796, p < .001]$ time windows. This is not surprising, given that both windows provide information about the types of trials being observed, and thus offer some redundancy with respect to the binary classification problem. Additionally, the relationship between classifier performance and individual ERP component amplitudes is presented in Figure 5.9b. In both cases, strong correlations suggest that the analyses successfully capture relevant single-trial features in the data. These results pose a trade-off with respect to the specificity of a particular classification analysis (i.e. the focus on specific cognitive processes and associated ERP components) and the overall performance that can be expected from a given analysis.

The three approaches used to evaluate class-relevant features in auditory mismatch data all tell a similar story: single-trial brain responses measured during standard and deviant trials are primarily distinguished by activity in time-windows corresponding to the MMN and P3a ERP components, with maximal differences at fronto-central electrodes. Additional brain activation in early (0-50 ms) and late (400-450 ms) time windows relative to stimulus onset are also informative. In essence, the classifiers trained in the present analyses make use of data features corresponding to a sequence of perceptual and cognitive processes triggered by a stimulus. Single-trial EEG measurements are modulated in a

consistent manner during deviant trials, providing the classifier with features that can be used to predict which type of trial has been observed.

## Estimating generalization performance

While the previous cross-validation methods are useful for estimating performance in the context of a particular classification problem, the ability to generalize to new datasets lies at the heart of real-world systems that use classifiers for different applications. This is referred to as 'generalization' or 'transfer learning' (van Gerven et al., 2009). Typically, generalization performance is lower than cross-validation performance due to non-stationary features that aren't adequately represented in the training data. In EEG data, many factors influencing the measurement of class-relevant brain signals will change between the collection of training data and the subsequent use of the trained classifier, such as the quality of electrode connectivity and the individual's subjective state (i.e. fatigue, concentration; see McFarland and Wolpaw, 2005; Blankertz et al., 2010). These effects are more pronounced in paradigms that make use of across-session designs in which training data is collected on one day, followed by online use of the classifier on subsequent days. Therefore, it is useful to estimate performance in these settings.

Figure 5.10 shows the mean generalization performance obtained for individual participants using Dataset 2, for both within- and across-session designs. For the within-session data, performance decreases significantly in the second $[t(11) = 4.569, p < .001]$ and third measurement blocks $[t(11) = 4.222., p = .0014]$ relative to the cross-validation rates obtained when training a classifier using data from the first block, following a similar pattern as seen with the cross-validation rates obtained when training classifiers using these two blocks (shown in Figure 5.5). The mean across-session generalization performance was also significantly lower for the second $[t(11) = 4.340, p = .0012]$ and third $[t(11) = 5.391, p < .001]$ sessions than the cross-validation rates obtained when training a classifier using all available data from measurements on the first session. Here, the drop in performance is larger, given that the initial cross-validation rates are also higher than when training using only data from the first measurement block. However, in both the within- and across-session versions of the analysis, the average generalization rates are still significantly above chance level, with performance approximately between 60 and 65%.

Another form of generalization that has drawn interest is the ability of classifiers to generalize across participants. Here, a classifier is trained using data from multiple individuals, and is then tested using another individual's previously unseen data. Such analyses have been used to investigate the extent to which various forms of brain activity associated with auditory perception and language processing overlap between individuals (Schaefer et al., 2011; Herrmann et al., 2012; see also Chapter Three). The use of cross-participant classifiers has several potential advantages: firstly, the use of data from multiple participants provides a means for assembling relatively large datasets for classifier training, which, as was discussed previously, can improve performance. Additionally, the use of a previously trained cross-participant classifier could eliminate the need for collecting training data during online applications, thereby reducing the amount of time required along with associated issues such as response habituation and fatigue. On the other hand, cross-participant classifiers might also fail to capture individual variability in class relevant features.

The results of cross-participant analyses of both Dataset 1 and Dataset 2 are presented

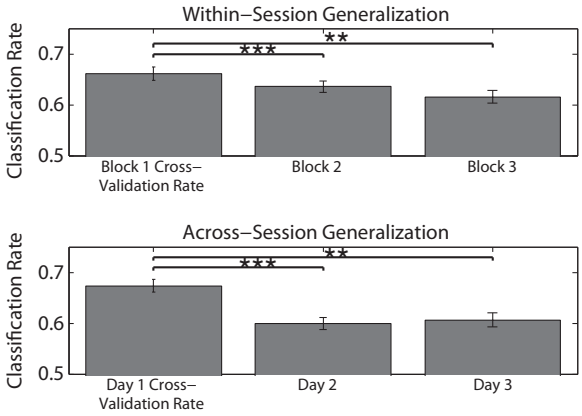


Figure 5.10: *Within-participant generalization across measurement blocks and sessions. The ability of classifiers trained on individual data to generalize to novel data collected during either the same or in different measurement sessions was evaluated using Dataset 2. In the within-session analysis, classifiers were trained using data collected in the first measurement block and then tested using data from the second and third blocks. In the across-session analysis, classifiers were trained using data from all three blocks measured in the first session, and tested using the initial blocks measured in the second and third sessions. For both analyses, a significant drop in performance was observed relative to the cross-validation rates obtained when training the classifiers. The significance of these paired-sample t-test comparisons is indicated using asterisks:* $** = p < .01$, $*** = p < .001$. *The drop across sessions was more severe, and could be potentially due to a number of factors, including changes related to the placement of electrodes as well as differences in the subjective states of the individual participants on different days.*

in Figure 5.11a. Here, a comparison is presented of the cross-validation rates obtained when training a classifier using individual data and the performance obtained when applying a classifier trained using data from all other participants (within the same dataset) to the same individual data. The analysis for Dataset 1 made use of all available data, while the analysis of Dataset 2 made use of 3 blocks of data (maximally 1800 trials per participant). On average, there was approximately a 2-3% decrease in performance relative to the individual cross-validation rates. This decrease was significant for both Dataset 1 [$t(13) = 3.876, p < .01$] and Dataset 2 [$t(11) = 5.056, p < .001$]. This decrease is substantially less than the across-session generalization performance that was seen in the analysis of individual data from Dataset 2, suggesting that cross-participant classifiers might offer some advantages in paradigms making use of longitudinal measurements. Classification rates for all participants in both datasets were significantly above chance-level, based on the binomial confidence intervals. In order to verify that that these results were not influenced by a bias in the data, a permutation test of the results was also conducted by shuffling the labels of the individual epochs (10000 permutations) and estimating the p-value as the fraction of the permutation distribution with scores greater than or equal to the observed classification rate. Here, all of the results were highly significant ($p < .001$, except for PP10 in Dataset 1, $p = .0025$).

An additional question regards the extent to which these classifiers generalize to novel datasets collected using different stimuli and measurement paradigms. This was evaluated by applying the cross-participant classifiers trained in the previous analysis (using all participants) to the other dataset. The results can be seen in Figure 5.11b. While significant decreases in performance relative to the cross-participant generalization rates were observed for both Dataset 1 [$t(13) = 3.341, p < .01$] and Dataset 2 [$t(11) = 4.735, p < .001$], performance for both datasets was on average higher than the across-day generalization performance obtained when using individually trained classifiers. Only one participant in Dataset 1 (PP10, classification rate = .521) did not show classification rates significantly above chance level; classification rates for the remaining Dataset 1 participants (range: .566 - .734) and all Dataset 2 participants (range: .541 - .690) were significantly above chance. Such results suggest that the brain responses collected using different variations of the MMN paradigm and stimuli generalize well enough to obtain reasonably high classification accuracies. This in turn opens the door for online paradigms that make use of variable stimuli and sequence types, depending on required performance.

The average classification rates in the preceding analyses of generalization performance fall approximately between 60-66%. In BCI paradigms that make use of single-trial classification, it is common to combine predictions made for multiple data epochs in order to increase the accuracy of the predictions. Recalling equation 5.1 (see also the section 'Classification Analyses' in Chapter Three), classifier predictions can be combined by summing the decision values obtained for each epoch $x_i$ in a set of $k$ epochs.

Figure 5.11c presents an analysis of the improvements in accuracy obtained when combining single-trial predictions across differing numbers of trials. The analysis made use of the predictions obtained in the cross-participant and cross-dataset analyses of generalization performance. As can be seen, average classification rates rapidly improve when combining up to 10 trials, with the rate of improvement gradually decreasing thereafter. While the combination of multiple trials will lead to an increase in the accuracy of predictions, it comes with a trade-off: collecting multiple epochs of data requires additional time, and thus decreases the speed at which the system is able to make predictions. In the

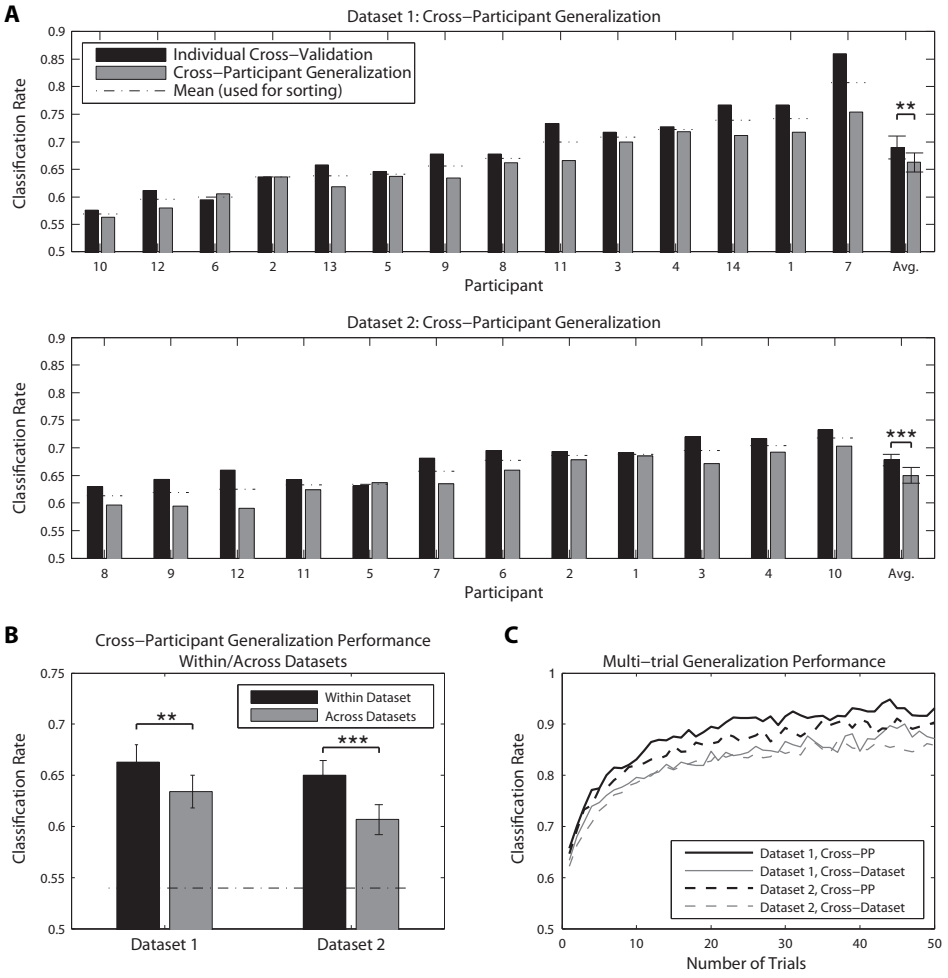


Figure 5.11: *Classifier generalization to novel participants and datasets. A) Both datasets were used to perform a cross-participant analysis of generalization performance, in which a series of classifiers were trained using all but one participant's data, and then tested on the remaining participant. The results of these analyses are presented along with the individual cross-validation rates for comparison. For both datasets, a significant decrease in performance relative to the individual cross-validation rates was observed (approx. 2-3% reduction). This drop in performance is most likely due to the variability of ERP features across individual participants, such as component latencies and spatial topographies. However, it also appears that these features generalize across participants to the extent that average classifier performance of 65% or higher is still possible. B) The ability to generalize across datasets was evaluated by training a classifier using all available data from one of the datasets, and testing it using the data from the other dataset. The average performance of these tests across individual participants is shown relative to the cross-participant classification rates obtained when training and testing within each dataset. As expected, a significant drop in performance was observed in both cases. However, in both cases, average classification rates were above 60%, well above the binomial confidence interval, and higher than the performance observed for within-participant generalization across sessions.* *Continues on next page...*

Figure 5.11: *continued. C) Classifier predictions can be combined across multiple trials to increase prediction accuracy. This was evaluated for both datasets using the results of the within-dataset cross-participant analysis and the across-dataset analysis. Results shown are averaged across all participants within each dataset. For panels B) and C), the significance of the statistical comparisons are indicated using asterisks:* $** = p < .01$, $*** = p < .001$

BCI literature, the term 'bit rate' is used to describe the informational output of a system (van Gerven et al., 2009), and is a function of the classification accuracy and the time require to obtain the data. Lower single-trial classification rates will lead to systems with a lower bit rate, due to either the number of incorrect predictions made or the increased amounts of time needed to collect multiple epochs in order to increase the accuracy of the predictions. The role of this trade-off between time and accuracy in the decoding of MMN responses for different applications is discussed in the subsequent sections.

In summary, the analyses above have shown that classifiers trained on individual and group MMN data generalize well to new datasets. While performance drops relative to the cross-validation rates obtained during classifier training, generalization performance within- and across-sessions, as well as across participants and data sets, is on average in the range of 60 to 65%, and can be improved by combining classifier predictions across multiple, non-overlapping trials.

## Online tracking of perceptual discrimination

The previous sections illustrated how classification analyses can be used together with single-trial data collected in an MMN paradigm. This section illustrates how these methods can be used to track the brain activity underlying the MMN and P3a responses online. In particular, it shows how the output of a logistic regression classifier can be interpreted as a probability that a particular type of brain activity (i.e. the MMN response) has been observed in a given epoch of data. Logistic regression classifiers make use of the logistic function, which always takes on values in the range $[0, 1]$:

$$p(+|x) = \frac{1}{1 - e^{-f(x)}} \tag{5.3}$$

where $p(+|x)$ is the posterior probability of the positive class given a data epoch $x$. A plot of the logistic function is shown in Figure 5.12a. Thus, the output can be interpreted as the posterior probability that the positive class has been observed by providing it as input to the logistic function. The probability of the negative class is equivalently $1 - p(+|x)$. Differences in the observed probabilities obtained when applying a classifier to a given data epoch reflect differences in the class-relevant features of the data. Thus, the observed probabilities can be used to order or index data on the basis of these features.

Figure 5.12b presents the distribution of decision values obtained for cross-validation test-set data in Dataset 1 during the cross-participant generalization analysis. Standard and deviant trials are plotted with separate colors. The corresponding probabilities for the trials can be inferred by referencing the logistic function plot above it. By dividing this data into subsets on the basis of ordered classifier decisions, it is possible to track underlying changes in the single-trial ERP morphology. Figure 5.12c displays ERP images

and ERP waveforms for 4 subsets of the data. No distinction between standard and deviant trials or participants is made. Rather, data are grouped on the basis of classifier decisions. Each subset of the data has a distinct morphology, with the primary differences related to the N1/MMN components, the P2/P3a components, and a late component between 400 and 450 ms following stimulus onset. For example, when comparing averaged data from the first quartile and the fourth quartile, the opposite patterns of activity can be seen at approximately 100 ms, 250 ms and 425 ms. Moreover, the gradual change in the average amplitude of the corresponding ERP components can be clearly observed in the second and third quartiles. This indicates that the decision values are tracking amplitude fluctuations and spatial shifts in brain activity at these specific points in time in the single-trial data. An alternative grouping of trials is presented in Figure 5.12d. Here, grand average ERPs are presented based on whether a trial was classified correctly. This corresponds to 'Hits' and 'Misses' for deviant trials, and to 'True Negatives' and 'False Positives' for standard trials. A similar pattern to the ERPs in Figure 5.12c emerges, suggesting that the trial-to-trial fluctuations in brain responses at specific time points within a given trial-type are being tracked by the classifier.

The ability to track fluctuations in specific ERP components at the single-trial level offers the possibility to investigate the short-term dynamics of the brain activity underlying their generation. The assumptions made by this approach are essentially the same as those upon which general ERP methodology is based. In essence, the time-locked brain activities measured in single-trial EEG data reflect meaningful sources of variation, such as individual differences (e.g. expert/non-expert), task differences (e.g. target/non-target trials), habituation effects or stimulus differences (e.g. standard/deviant trial). Using averaging, these differences in the time-locked ERPs are typically visible to the eye, and can be subjected to statistical analysis. However, such methods have no means of investigating trial-to-trial fluctuations in the generation of these response. Methods such as ERP images (Makeig et al., 2004) have been developed as a means of sorting data from individually measured trials post hoc using a variable of interest, such as oscillatory phase. Recently, models of MMN generation based on the free-energy principle have been used to explain trial-to-trial variation in MMN amplitudes (Lieder et al., 2013). The continuous output of a classifier trained on representative data provides another means of indexing single-trial brain responses that can also be used online.

Examples of classifier output across time can be found in Figure 5.13a. The results of the cross-participant analysis for 150 consecutive trials are used to visualize fluctuations in the continuous output of the classifier for both standard and deviant trials. Results are shown for three representative participants from Dataset 1. The effects of using different numbers of trials are also presented. Combining multiple trials leads to output which is generally more stable across consecutive predictions, but also reduces the frequency at which predictions are made about the data. Individual participants also show differences with respect to the relative strength of the predictions being made about standard and deviant trials, as well as the specific points in time when fluctuations in classifier output are observed. These differences are most clearly observable when combining across 10 trials.

Mean output across all participants in Dataset 1 is shown for different numbers of trials (1, 3, 5 or 10) in Figure 5.13b. In general, overall predictive confidence improves through the combination of multiple single-trial predictions. Continuous classifier output also tends towards higher values for standard trials than for deviant trials. This corresponds

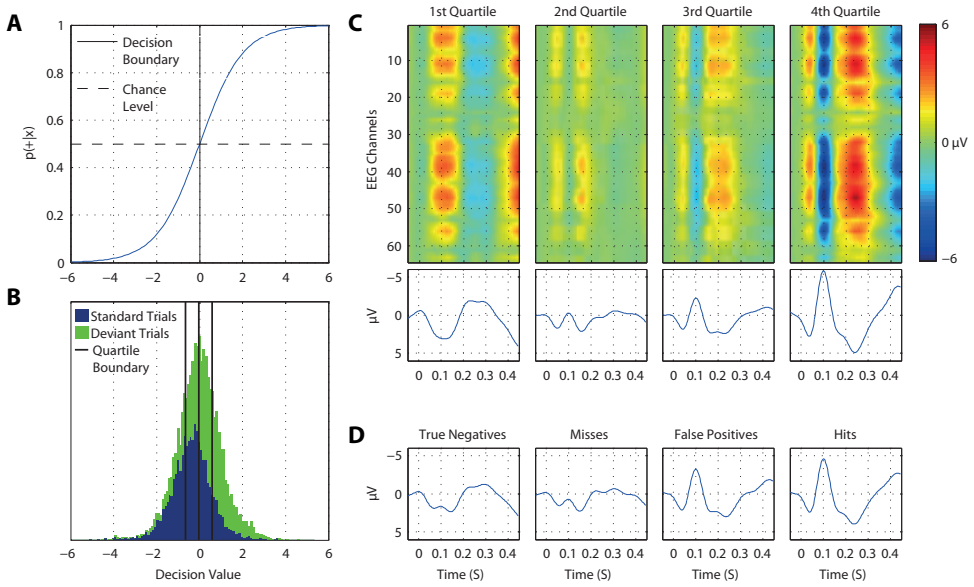


Figure 5.12: *Online tracking of single trial brain activity in an auditory mismatch paradigm. A) The logistic regression function. B) Sorted single-trial classifier decisions for test set data in the cross-participant analysis of Dataset 1 (see Figure 5.11a, upper panel). A division of the trials into quartiles on the basis of the decision values is shown using thick black lines. C) Grand average ERPs for sorted single-trial data in each quartile. Upper panels show ERP activations at all 64 EEG channels, while the lower panels show the averaged waveforms obtained at nine fronto-central electrodes (see Figure 5.1 for reference). A clear shift in both the number and amplitude of the components in the obtained ERPs is visible across the four subsets of the data. D) Grand average ERPs based on classifier signal detection performance. Correct and incorrect decisions for both the target (deviant) and non-target (standard) classes are used to group trials. As can be seen, shifts in the relative amplitudes of multiple ERP components are associated with correct and incorrect decisions for both standard and deviant trials.*

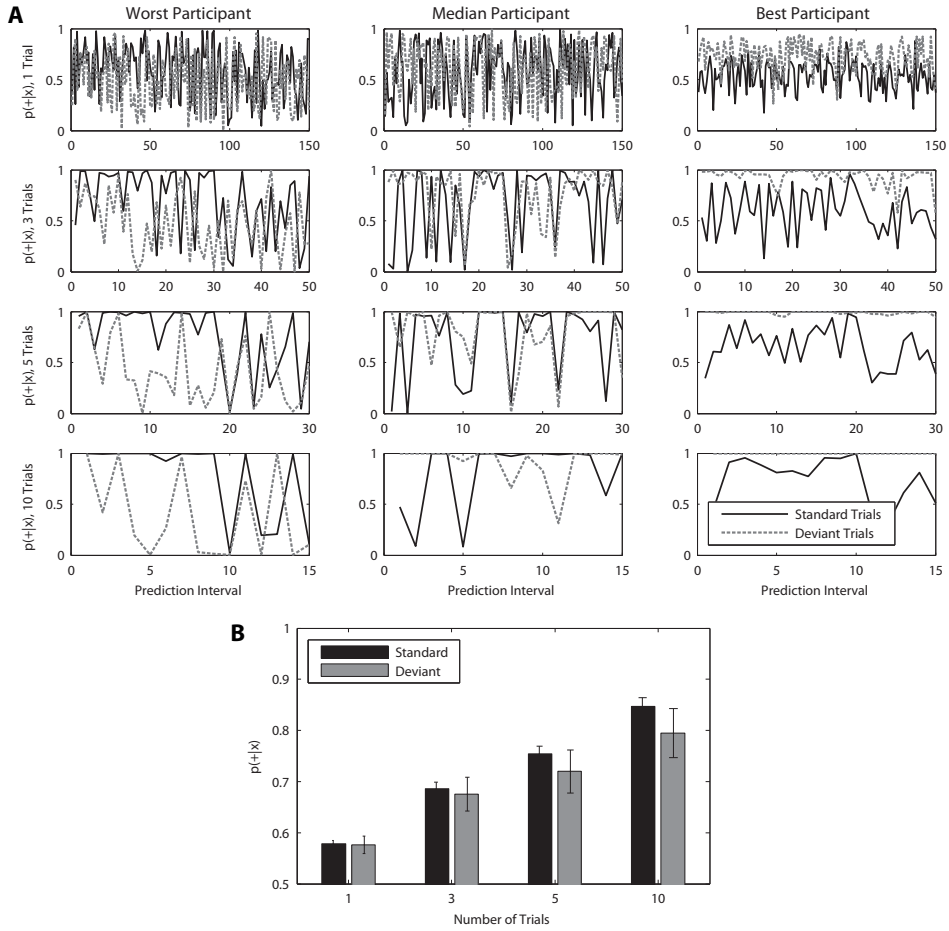


Figure 5.13: *Probabilistic interpretation of classifier output over trials. A) Data from three representative participants in Dataset 1 (left to right) illustrates a probabilistic interpretation of classifier decisions for single and varying group lengths of (non-overlapping) trials. Results are taken from the first 150 consecutive trials analyzed in the cross-participant analysis. From top to bottom, the number of trials used to make predictions about ongoing brain responses is varied, using either 1, 3, 5 or 10 trials. This effectively reduces the number of predictions that can be made: while 150 individual predictions are made when using only 1 trial, only 15 predictions are made when using 10 trials. As can be seen, the output becomes increasingly stable as the number of trials is increased. B) Similarly to the increase in single-trial classification rates observed when combining across multiple trials (see Figure 5.11c), the mean of the continuous classifier output for all 150 trials increases as predictions are combined across groups of trials of increasing size.*

to a slightly higher single-trial classification accuracy for standard trials (67.6%, see Figure 5.12d 'True Negatives') than deviant trials (64.9%, see Figure 5.12d, 'Hits'), as well as to a median decision value (see Figure 5.12b) that is slightly negative. While these differences are small, they suggest that classifier predictions about standard trials are more reliable and confident than those about deviant trials. In the context of the auditory mismatch paradigm being investigated, one possible explanation is that there is less variation in the brain responses measured on standard trials than in deviant trials, given the greater frequency (and predictability) of standard trials. In other words, brain responses measured in standard trials may be more stable as compared to deviant trials.

Given the average single-trial classification accuracies (approximately 66%) obtained in the present analysis, one factor that will influence classifier output is change in the EEG signal unrelated to the auditory evoked potentials being analyzed. This implies that the output of the classifier is noisy. While data can be combined across multiple trials to reduce the effects of this noise, this comes with a trade-off: rather than estimating the probability of a particular set of brain responses at the single-trial level, an estimation of these responses is made across a longer period of time. This means that changes in the brain responses across shorter intervals will be mixed together, and that a delay is introduced into the tracking procedure.

The present approach of tracking brain responses underlying perceptual discrimination is in many ways similar to BCI paradigms which also make use of single-trial pattern classification analysis. A key aspect is the interpretation of classifier output as a continuous probability rather than as a binary decision. These values correspond to graded modulations of brain activity measured using EEG at time points where ERP components such as the MMN and P3a are typically observed, and as such, serve as an index of ongoing perceptual discrimination. The ongoing fluctuations in classifier output effectively track changes in perception online. Differing degrees of single-trial classification accuracy for standard and deviant trials are also reflected in the continuous interpretation of classifier output, suggesting that the patterns of brain activity measured in individual standard trials are somewhat more stable than in deviant trials. While the accuracy of this tracking procedure is also influenced by noise sources in the EEG signal, predictions can be combined across multiple trials in order to obtain a more stable measure of perceptual discrimination across longer time intervals.

# Potential domains of application

This chapter has thus far presented a method for utilizing pattern classification methods and single-trial EEG data recorded in two variants of the auditory mismatch paradigm to track brain activity associated with perceptual discrimination processes. It serves as a 'recipe' for online applications that aim to monitor changes in perceptual discrimination, including the effects of learning.

## Cognitive Monitoring and Passive BCI

One domain where these decoding methods can be applied is cognitive monitoring, in which real-time measurements are used by clinicians and researchers to infer the cognitive state of a user. For example, the working memory load of an individual can be monitored

as they engage in challenging tasks using single-trial decoding of EEG data (Brouwer et al., 2012). The ability to detect ongoing changes in working memory and cognitive load has also been proposed for use in the enhancement of aspects of human-computer interfaces (Grimes et al., 2008).

Variations in MMN response characteristics are associated with both differences in perceptual discrimination abilities and with various clinical and medical conditions. The ability to decode MMN responses at the single trial using existing classifiers and to perform a probabilistic assessment of the MMN response using a reduced number of trials and/or electrode recording channels might offer a valuable alternative in settings where assessment of brain responses augments or is preferred to behavioral assessments, but in which time is limited. This might include forms of objective audiometry (Bukard et al., 2007), which measure perceptual thresholds using auditory evoked potentials instead of behavior.

Closely related to cognitive monitoring are passive BCIs (Zander and Kothe, 2011), which measure ongoing brain responses as an auxiliary input to an interface. For instance, a computer could automatically adapt the difficulty of a given learning task on the basis of ongoing measurements of working memory load. Similarly, a passive BCI based on an auditory mismatch paradigm would be able to adapt the difficulty of an auditory learning task on the basis of a probabilistic assessment that an MMN response had been observed in the preceding trial(s). Additional work should assess how adapting the difficulty of a listening task (i.e. stimulus contrast size) influences classifier performance, as changes in contrast salience are known to influence the timing and amplitude of the MMN component (Näätänen et al., 2007).

## Neurofeedback

Neurofeedback is an approach in which measurements of an individual's ongoing brain activity are mapped onto a feedback signal of some kind, typically in the auditory or visual modality (Hammond, 2011). The goal of neurofeedback is to modulate the targeted forms of brain activity, which are associated with particular cognitive or mental states, such as attention, concentration or anxiety (Lansbergen et al., 2011; Zotev et al., 2011). For instance, many EEG-based approaches use measurements in the frequency domain of oscillatory activity at specific electrode locations in the alpha, theta and other bandwidths (Hammond, 2011). Neurofeedback paradigms using fMRI measurements can measure activity in specific brain regions, such as the limbic system (Zotev et al., 2011) or perceptual cortices (Yoo et al., 2006; Scharnowski et al., 2012). Neurofeedback approaches have shown promise in treating various clinical disorders such as chronic pain (de Charms, 2008) and tinnitus (Weisz et al., 2011), and can also improve cognitive performance (Zoefel et al., 2011) and emotional regulation abilities (Johnston et al., 2010).

Two studies recently demonstrated that neurofeedback training using fMRI measurements of activity in early visual cortex can induce perceptual learning effects (Shibata et al., 2011; Scharnowski et al., 2012). These findings are relevant for the present method, as modulations of the MMN response have also been linked to perceptual learning (Tremblay et al., 1997; Menning et al., 2000). In particular, Shibata et al. (2011) made use of neurofeedback based on the decoding of specific activity patterns in visual cortex corresponding to three Gabor stimuli with different orientations. This was done using a logistic regression classifier. Similarly to the present method, the neurofeedback was based on a probabilistic interpretation of a classifier's output when applied to novel data during the

neurofeedback sessions. Their results indicated that the mean probability that the targeted class of brain response had been observed (i.e. the feedback signal) increased on average across the course of the training sessions. This suggests that the continuous changes in classifier output observed online and provided as feedback reflect meaningful variations in the decoded brain activations, and that modulation of this feedback signal can lead to concomitant perceptual learning effects.

Given the link between MMN response characteristics and the perceptual skills underlying language and music proficiency, neurofeedback based on single-trial decoding of MMN responses has the potential to augment language and music training paradigms. Future research into MMN-based neurofeedback paradigms should investigate the specific task and stimulus parameters that optimize the effectiveness of the neurofeedback training paradigm. This might include issues such as stimulus contrast size, task instructions (active or passive use of the neurofeedback), and the number of trials used to generate feedback. The trade-off between accuracy (i.e. single-trial classification rates) and time (i.e. number of trials used to make predictions) is an issue that has been considered in the context of reinforcement learning (Cardinal, 2006). As both temporal delays and noise influence the effectiveness of a reinforcer stimulus (in this case, the classifier decisions mapped onto a feedback signal), either the use of single- (noisy) or multi-trial (delayed) classifier output will reduce the effectiveness of the feedback. It is worth noting that the mean classifier output in the study of Shibata et al. (2011), which was used to induce perceptual learning via neurofeedback training, was comparable to that obtained here. This suggests that even noisy neurofeedback signals can be effective in modulating targeted forms of brain activity.

# Conclusion

Pattern classification analyses play an increasing role in fundamental and applied research settings that make use of neuroimaging. This chapter has presented a framework for decoding analysis of EEG data collected in an auditory mismatch paradigm, along with guidelines for the development and optimization of online paradigms based on this framework. Firstly, it demonstrated how MMN data collection and preprocessing can be optimized for pattern classification analysis through careful selection of filter frequencies and appropriate temporal downsampling. The parameters used in these steps were shown to generalize across MMN datasets. It then showed how searchlight, AUC and ERP based methods can be used to evaluate the contribution of components such as the MMN and P3a to single-trial decoding performance. Additionally, estimates of both across-session and across-individual generalization performance were presented. Finally, it was shown how the same decoding methods can be used online to index fluctuations in the amplitude of brain responses at different points in time in single-trial data. This framework can serve as a basis for subsequent research seeking to implement specific online applications based on an auditory mismatch paradigm.

# 6

# General Discussion and Conclusions

The past four chapters presented the results of research conducted during the development and initial testing of a neurofeedback paradigm for auditory perceptual learning. Before discussing both the challenges and future directions for this line of research, I first briefly summarize what has been presented.

Chapter Two contains the results of an ERP study that investigated differences in the perception of English phonemes by native and non-native speakers of English using a mismatch negativity (MMN) paradigm. The non-native speakers all spoke Dutch as a native language, and were highly proficient speakers of English. The stimuli in the experiment were steps on an English /pa/-/ba/ continuum of consonant-vowel syllables, and were chosen on the basis of specific differences in English and Dutch phonology. Specifically, English makes use of aspiration as an acoustic cue for distinguishing voiced and voiceless stop consonants, while Dutch stop consonants are unaspirated, and instead make use of pre-voicing to distinguish the voiceless and voiced versions. Thus, while both languages make use of voice-onset time (VOT) to distinguish the two categories, the specific acoustic cue that is used by the two languages differs. Results of a behavioral identification task revealed a different category boundary for the native and non-native speakers, with the category boundary for native-Dutch speakers occurring at a reduced VOT relative to native-English speakers. This behavioral difference corresponded to a group-level difference in the amplitudes of the MMN component of ERPs measured when using both a within-category and a category-boundary (native) stimulus as deviants. This implies that the perceptual sensitivity of the two groups to aspiration cues differs as a function of native language, despite the extensive second language experience of the native-Dutch speakers.

This same dataset was then used in Chapter Three to investigate the use of multivariate pattern classification analysis in conjunction with single-trial EEG measurements made in the MMN paradigm. Here, we aimed to determine the reliability with which so-called decoding analysis could be used within-participant to distinguish individual brain responses (i.e. within-participant) measured on standard versus deviant trials (i.e. those containing the MMN response), and the extent to which this is influenced by differences in stimulus parameters and native language. An additional set of classification analyses investigated

the use of cross-participant datasets on the reliability of decoding performance. Results indicated that, on average, classification rates were higher across the different stimulus conditions for native-English speakers, particularly in the case of the within-category condition. At an individual level, the performance of the decoding analyses across stimulus conditions showed strong relationships with both ERP features (i.e. mean MMN amplitude) and behavioral identification scores. Reliability also improved on average when combining classifier predictions across multiple trials. Interestingly, the results of the analysis of the cross-participant datasets indicated that the overall performance of the decoding analyses improved on average with the inclusion of data from multiple participants. In particular, a large improvement in average classifier performance was observed relative to the within-participant analyses for the native-English speakers when training on the within-group dataset. However, when applying a cross-participant classifier trained using data from the native-English speaking group to data from the native-Dutch participants, a drop in average performance was observed relative to within-participant results. These results suggest that, while the pattern of single-trial brain responses to speech stimuli generalizes across individuals, these patterns differ across native-language groups. This in turn implies that individual differences in single-trial measurements of brain activity underlying speech perception can be reliably captured using pattern classification analysis techniques, opening the door for brain-computer interface and neurofeedback paradigms based on speech perception.

The results of an initial experiment investigating the effects of a decoded-EEG neurofeedback paradigm on both brain responses and perceptual learning are reported in Chapter Four. This neurofeedback was based on the output of a logistic regression classifier trained on data collected in a mismatch negativity paradigm. The experiment made use of simple sinusoidal tone stimuli with different fundamental frequencies, and investigated changes in behavioraly measured frequency discrimination sensitivity across the course of four experimental sessions on different days within one week. The tone stimuli were preferred to speech stimuli given the well-established behavioral and electrophysiological changes associated with perceptual learning in this domain (Micheyl et al., 2006; Menning et al., 2000). Two groups of participants were used in the study: an experimental group that received neurofeedback training during each of the four sessions, and a control group that received a form of sham feedback. During an initial measurement made prior to the neurofeedback/sham-feedback training sessions, both groups showed similar patterns of results for both behavioral and EEG measurements. However, during the feedback sessions, the MMN response amplitudes of the experimental group were enhanced relative to those of the control group, suggesting that the neurofeedback had effectively modulated the targeted brain response. With respect to the behavioral measurements of frequency discrimination sensitivity, both groups showed consistent improvements in performance across the four days of the experiment, but no differences between the two groups were observed. Thus, while the neurofeedback procedure did not lead to an enhancement of perceptual learning, it effectively modulated MMN response amplitudes during the feedback sessions.

Chapter Five presented an overview of the methods used for pattern classification analysis of EEG data collected using an MMN paradigm. Two datasets collected using different variants of the MMN paradigm served to illustrate various aspects of the classification approach. In particular, it focused on methods for optimizing both data preprocessing and collection methods to obtain maximal classification performance, including an assessment of the effects of steps such as filtering and temporal downsampling as well as dataset size

on the obtained classification rates. It also presented methods for evaluating data features that contribute to overall classification accuracy, estimating generalization performance to novel datasets and methods for utilizing continuous classifier output in the context of online applications such as cognitive monitoring, passive BCI and neurofeedback. The recommendations provided by the chapter are intended for use as a general-purpose framework for future studies that make use of decoding methods in conjunction with data collected in an MMN paradigm.

This concludes the overview of the dissertation. In the following sections, I discuss the present state of research into applications of neurofeedback for perceptual learning, and highlight some of the particular challenges faced by approaches based on measurements of the MMN response. Afterwards, I discuss several potential steps that can be taken to address these challenges before considering more general issues such as the ethical implications of the present research and its potential impact in educational settings.

## Decoded-EEG Neurofeedback for Auditory Perceptual Learning: the present

In the time since this project was started, interest in the use of decoding methods for single-trial analysis of the MMN response (Herrmann et al., 2012; Tzovara et al., 2013), and in neurofeedback paradigms geared towards perceptual learning (Shibata et al., 2011; Scharnowski et al., 2012) have grown substantially. Crucially, two studies carried out using fMRI and neurofeedback paradigms based on visual perception have demonstrated an enhancement of the targeted visual perception skills in participants who underwent neurofeedback training based on measurements of brain responses in primary sensory cortex. One of these studies (Shibata et al., 2011) made use of a decoding approach based on a logistic regression method very similar to the one developed in Chapter Five and used in the experiment presented in Chapter Four. This suggests that this method represents a domain general approach that can be utilized in applications, such as neurofeedback and cognitive monitoring (Zander and Kothe, 2011), which rely on real-time neuroimaging.

Here at the Donders Institute, the EarOpener project has expanded upon the themes explored in this project, and has begun to develop additional variants of the neurofeedback paradigm intended for language learning. Difficult to learn phonetic contrasts in Japanese, English and Chinese have been incorporated into different types of tasks in order to further advance our understanding of the cognitive processes involved in this type of neurofeedback loop, and to improve the effectiveness of the feedback itself. While no increased efficiency of perceptual learning has been observed in the current paradigm yet, the ability to modulate the underlying brain responses suggests that further refinements to the paradigm might very well lead to such effects. Subsequent sections of this chapter discuss several important aspects of the neurofeedback paradigm, such as task, stimuli and training data collection, along with potential refinements that could improve its effectiveness.

## Cognitive mechanisms underlying neurofeedback use

One issue that remains unresolved is the role of the task given to users of the neurofeedback system, and the distinct cognitive mechanisms that are engaged by different instructions.

Chapter Four presented results of an experiment in which the participants were blind to the neurofeedback procedure. Instead, participants were instructed to attend to the films they were presented in order to answer questions about their contents. While a modulation of the MMN response relative to control participants was observed, the fact that participants were unaware of the goal of the experiment might also partially explain why no effect of neurofeedback was observed in the frequency discrimination threshold measurements. In a recent review of the evidence for both cognitive and associative accounts of learning, it was concluded that robust and replicable instances of unconscious learning have failed to emerge in the experimental literature (Shanks, 2010).

Recent findings suggest, however, that implicit learning of statistical regularities in our environment can influence our perception (Kaufman et al., 2010; Turk-Browne et al., 2010), specifically in the case of language learning (Conway et al., 2010). There is also evidence that artificial modulation of brain activity in language-related brain regions using trans-cranial direct current stimulation (tDCS) leads to an enhancement of implicit learning effects in language tasks (de Vries et al., 2010). This suggests that further refinements of the neurofeedback protocol used in Chapter Four might very well lead to perceptual learning effects. Whether or not an appropriately designed paradigm making use of an active learning and neurofeedback task would enhance perceptual learning remains an empirical question.

The current choice to make use of a design in which participants were blind to the neurofeedback procedure was based in part on the results of a previous pilot experiment that involved an active task instruction (see Appendix B). Here, participants were specifically told that the neurofeedback was based on their brain's response to the presented sounds, and that they should actively try to modulate the feedback signal (i.e. try to increase the clarity of the film). This led to an interesting result: while no clear modulation of the MMN response was observed, a progressive increase in the amplitude of the P3a response was observed across participants.

The P3a response is associated with a frontal generator distinct from the bilateral generators of the MMN response in auditory cortex, and is considered to reflect a shift of attention following the presentation of a perceptually salient deviant stimulus (Näätänen et al., 2007). This component is distinguished from the P3b response, a larger component which overlaps in time with the P3a components. Together, these two subcomponents are observed as the classical P300 response (Polich, 2007), and are elicited in active oddball tasks such as those employed in various BCI systems. (Farwell and Donchin, 1988; Sellers and Donchin, 2006; Schreuder et al., 2010; van der Waal et al., 2012). Given the disassociation of the P3a and P3b components in different task settings, researchers have concluded that the elicitation of the MNN response does not require attention, even though its amplitude can be modulated in some cases by attention (Sussman, 2007).

Recent attempts have also been made to interpret the MMN response with respect to the predictive coding framework (Garrido et al., 2009; Lieder et al., 2013) and in the broader context of cognitive models of auditory scene analysis (Winkler et al., 2009). Both interpretations focus on the role which top-down expectations regarding the auditory environment play in generating predictions about incoming sensory information. Unexpected events lead to the generation of a mismatch response, indicating a prediction error that leads to an updating of our subsequent predictions about the environment. Crucially, both these accounts of suggest a potential role for top-down influences on the generation of the MMN response, independent of attention. This in turn implies that alternative strategies

to the purely passive or active instructions employed in this thesis may be possible.

The interactions between task instruction, attention and the MMN response also highlight another unresolved issue: the role of instructions during the collection of training data for the classifiers utilized during the neurofeedback session. It is well established that active instructions (e.g. attend to the deviant stimuli) will lead to the elicitation of a P300 response whereas a passive instruction will not (Sams et al., 1985). However, a difference in instructions between measurements used to collect training data and those used during neurofeedback could also lead to changes in the relative timing, amplitude and number of components elicited in the two types of ERPs observed. This in turn could lead to a reduction in classifier performance during neurofeedback due to the non-overlapping features of the two datasets. Further investigation of the ERPs elicited under different instruction conditions is needed in order to determine the optimal method for providing instructions during training data collection and actual neurofeedback use.

## Technical Challenges

One of the big challenges still faced by EEG-based BCI systems is the reliability of single-trial classifier performance. While developments over the last decade in preprocessing, feature selection and classification methods have all led to improvements in the performance of non-invasive BCI systems, average classification rates across users are typically less than 80% (i.e. 1 out of 5 predictions is incorrect, see Sellers and Donchin (2006); Nijboer et al. (2008); Schreuder et al. (2010); Halder et al. (2010)). For data collected using an MMN paradigm, the single-trial performance is even lower, typically between 60-70% on average for perceptually salient stimulus contrasts, with rates decreasing as the salience of the contrast is reduced (see Chapter Three). The low performance (i.e. $<$60%) of classifiers trained using data collected with difficult to perceive stimulus contrasts, such as those representing contrasts at or near the behavioral discrimination threshold, furthers constrains the use of challenging materials.

While the accuracy of BCIs can be improved by making predictions on the basis of multiple-trials, this leads to a reduction in the overall speed with which the system can be used to generate output, referred to as the bit-rate (van Gerven et al., 2009). In the case of our neurofeedback approach, this translates into a delay between the events upon which feedback is based (i.e. the brain response to a sound event) and the presentation of the feedback. Research investigating the effects of noisy feedback and temporal delays on reinforcement learning suggests that increases in noise and delay progressively reduce the effectiveness of reinforcer stimuli (Cardinal, 2006). As the neurofeedback signal is a type of reinforcement stimulus, this implies that either delays or noise (i.e. incorrect feedback) will reduce the effectiveness of the neurofeedback. Future research should investigate both the impact of single-trial classification accuracy and temporal delays on the effectiveness of neurofeedback training.

Another challenge in the present paradigm is the role of response habituation. Typically, the amplitude of components such as the N1 and MMN diminishes following initial measurements when repetitive stimulation is employed, such as is the case with the oddball sequences used in the MMN paradigm. In paradigms making use of classifiers trained on individual participant's data, such as the one used in Chapter Four, this can lead to a situation in which the responses measured during the presentation of neurofeedback are

habituated relative to the training data. Moreover, depending on the type of stimulus contrast employed as well as individual differences in response features such as component amplitude, differing amounts of training data are required in order to obtain optimal single-trial classification performance and generalization to novel data. This could imply the need for longer measurements in order to collect a sufficient amount of training data, with these longer measurements then being subject to habituation effects.

In the case of complex stimuli, such as speech and musical sounds possessing dynamic spectral and temporal envelopes, classification analysis is further complicated by the relative shifts in component amplitudes and latencies that come about due to timing differences in their principal acoustic features. Such relative differences in component latencies and amplitudes provide discriminative information in the context of a classification analysis that are not necessarily related to the brain responses of interest; namely, those related to perceptual discrimination, such as the MMN. In the ERP literature, such differences are typically dealt with by measuring the same sounds in different perceptual contexts (i.e. as both standard and deviant stimuli in an MMN paradigm). While this is also possible in the present approach, it has the undesirable consequence of increasing the amount of training data that is required, thus introducing additional time requirements prior to the actual presentation of neurofeedback.

Individual differences in the reliability of the MMN response as an index of perceptual discrimination ability also pose a challenge for the neurofeedback approach defined in this thesis. In a statistical analysis of individual MMN responses collected using a highly salient stimulus contrast, it was found that approximately 80% of participants showed a clear MMN, even though all of these participants had normal hearing and showed good behavioral discrimination abilities (Bishop, 2009). This implies that 1 out of 5 individuals might be unable to make use of a neurofeedback system based on measurements of the MMN. In the broader context of BCI research, there has also been a discussion of 'BCI Illiteracy', or the fact that not everyone is able to successfully control a given BCI system (Vidaurre and Blankertz, 2010). Such differences can be due to a number of factors, including individual differences in brain morphology and other physiological factors that influence the propagation of EEG signals (Polich, 1989). A fully developed version of the present paradigm requires a means of assessing the suitability of individual users prior to extensive use of the neurofeedback system.

## Development of the Paradigm

The previous two sections highlighted some of the unresolved questions regarding the design of a decoded-EEG neurofeedback paradigm for perceptual learning applications, as well several technical challenges. This section proposes several potential steps that could be taken to address these issues and further develop the present paradigm.

Chapters Three and Five both presented the results of cross-participant classification analyses, in which the ability of a classifier to generalize across individuals was assessed. In most cases, these analyses led to similar performance, with some of the cross-participant analyses in Chapter Three actually showing higher average performance than the individual analyses. Although this may have been due to the larger size of the training dataset that was available in the cross-participant analysis, this is nonetheless an important observation when considering potential improvements to the design of the neurofeedback

paradigm. While most of the BCI systems reported in the literature make use of individually trained classifiers, as did the experiment presented in Chapter Four, a move towards cross-participant classifiers would eliminate the need for the collection of individual training data, thus reducing the amount of time needed during neurofeedback sessions, as well as the habituation effects observed in Chapter Four. However, the possibility also exists that an overall drop in performance of cross-participant classifiers relative to classifiers trained on individual data would outweigh these benefits, and remains to be determined empirically.

A vital step in the development of the current paradigm will be an evaluation of the role that it can play in existing educational settings related to language and music learning. For instance, are there specific advantages or disadvantages of a neurofeedback paradigm for perceptual learning relative to the perceptual learning that occurs in typical educational settings such as language and music courses? Or is there a particular time window in the course of perceptual learning during which neurofeedback would be most effective? While previous results on the time course of perceptual learning suggest that the brain responses underlying perceptual discrimination might arise prior to changes in behavioral discrimination abilities (Tremblay et al., 1998), it remains to be seen whether or not the introduction of neurofeedback training at this stage of learning leads to a relative enhancement of an individual's ability to perceive novel sound contrasts.

Finally, it is also worth noting that current developments in the technologies utilized in the neurofeedback paradigm will also directly contribute to improvements in its effectiveness and usability. For example, there is an increasing availability of cost-effective and portable EEG systems which offer features such as wireless data transmission and saline-based (as opposed to gel-based) electrode placement methods, both of which reduce the amount of time required for cap-fitting. Additionally, the increasing availability of powerful, portable computing resources (i.e. tablet computers) allows for the development of applications implementing the present paradigm that could be used outside the laboratory. Taken together, such developments imply that it would be feasible to bring a more developed version of the present training paradigm into settings such as the home or the classroom.

## Ethical Questions

Together with the powerful insights into the brain's functions that have been gained with recent developments in the cognitive neurosciences comes an important philosophical debate over the ethics of these methods. The field of neuroethics is host to wide-ranging discussions on the issues and potential implications surrounding topics such as BCI and neurofeedback (Haselager et al., 2009; Vlek et al., 2012). Topics addressed by these discussions include individual privacy, elective enhancement and mind control. Consider the following hypothetical situation: suppose it were possible, for instance, to determine whether a person has a predisposition to commit violent crimes on the basis of the results of some kind of brain-reading analysis that can be performed during childhood. The availability of such a technology itself would certainly lead to a debate over whether such information is private, and whether or not its use could be mandated by society.

Similar issues can be considered with regards to the present approach: could brain data that reflects ones perceptual discrimination abilities be used to either positively or

negatively profile someone in an educational or clinical context? If so, what are the consequences of the widespread use of such technology? Would the potential advantages of improved educational and/or clinical practice outweigh such consequences? Take for example the finding from Chapter Three, in which it was shown that the native language of a given participant could be predicted with above 80% accuracy on the basis of brain data. On the one hand, such a method could complement behavioral evaluations in the context of language training programs. On the other hand, although seemingly far-fetched, it could also be used to negatively profile someone in situations of conflict or war; for instance, a story exists that, in the second world war, a tongue-twister with the word 'Scheveningen' was used to distinguish Dutch and German individuals, as Germans weren't able to correctly pronounce 'sch'.

Consider another potential situation also closely related to the neurofeedback paradigm explored in this dissertation. What if a technology were available that allowed for all types of cognitive enhancements, such as improved perceptual and working memory function, but this technology was expensive and only individuals with high socioeconomic status could afford it (e.g a method that required fMRI measurements). The exclusivity of such technology might serve to reinforce socioeconomic stratification. While this would most likely not be the case for the current approach, it is worth keeping mind as methods based on expensive neuroimaging methods continue to develop.

In any case, the neurofeedback technology discussed in this dissertation would need to be highly reliable in order for its use to be practical. A problem faced by BCI and neurofeedback researchers is the gap between what is possible with current methods and technology, and the 'hype' with which such research topics are presented in the mainstream media. General audiences reading reports of developments in these areas of research often come away with unrealistic ideas about what can be achieved. As discussed earlier, the single-trial classification rates observed across the studies presented in this dissertation can be improved upon, meaning that the applications of the neurofeedback method would need to take potential concerns about the accuracy of the feedback into account. Moreover, the fact that not all individuals show a clear MMN response, even though they are able to behaviorally discriminate the same stimuli (Bishop and Hardiman, 2010), also limits the range of applicability of methods based on the decoding of the MMN response.

## Concluding Remarks

The research presented in this dissertation has focused on the development of a method for tracking changes in single-trial measurements of brain responses to sound contrasts, such as those encountered in language and music. This method in turn was used as the basis of a training paradigm that provides individuals with ongoing neurofeedback on these same brain responses. Crucially, it has shown that not only can these changes in brain responses be tracked using pattern classification techniques similar to those employed in EEG-based brain computer interfaces, but that these same brain responses can be modulated by the presentation of neurofeedback. Further continuation of this work will aim at improvements of the neurofeedback procedure that eventually lead to the development of training paradigms that can be used to augment or enhance auditory perceptual learning in the context of language and music training.

While this research is still ongoing, and the methods are still being refined, they rep-

resent a small part of a bigger transition in neuroscience research towards measuring the brain's activity in real-time. The past decades have already seen a dramatic increase in the understanding of brain function, and in the overall amount of basic and applied neuroscience research being conducted across the globe. We are now moving towards techniques that allow deciphering of the mind's contents on the basis of brain measurement alone, and which allow us to better understand the complex abilities that make us human.

# Appendix A

# Supplementary materials for Chapter Five

The two datasets analyzed in Chapter Five were collected in two pilot experiments that implemented a neurofeedback loop based on the decoding of single-trial EEG data collected using a mismatch negativity (MMN) paradigm, in which participants view silent film clips while passively listening to auditory sequences containing standard and deviant stimuli. The first experiment (Dataset 1) was conducted on a single day, and made use of oddball sequences typically associated with MMN recordings (Näätänen et al., 2007). The second experiment was conducted over four days and made use of so-called optimal MMN sequences containing 5 different types of deviant stimuli (Näätänen et al., 2004, 2007).

Information specific to each of the two experimental designs is reported before then presenting acquisition and data processing methods common to both of the datasets.

## Dataset 1: Participant information and design

### Participants

14 participants (aged 18-56) reporting normal hearing and with normal or corrected-to-normal vision participated in the experiment. The participants were recruited through either the Radboud University Research participant system or through the Department of Cognitive Artificial Intelligence at the Donders Centre for Cognition. All participants provided informed consent prior to the experiment. The research was conducted with the approval of the ethics committee of the Faculty of Social Sciences at the Radboud University Nijmegen.

### Procedure

Participants completed an initial measurement of their frequency discrimination threshold at 500 Hz using a two-alternative forced choice (2AFC) staircase procedure (3-up, 1-down, estimating the 79% correct threshold, Levitt 1971). Participants then completed four blocks of EEG measurements, the data from which were used to train a quadratically

regularized logistic regression classifier that was utilized in the subsequent neurofeedback blocks. Three of these blocks made use of oddball sequences containing a 500 Hz standard stimulus and a deviant stimulus set to the individual participant's frequency discrimination threshold. The final block presented only repetitions of the stimulus at the individual threshold (i.e. no deviants). Prior to the neurofeedback blocks, participants completed another frequency discrimination threshold measurement. Four additional EEG measurement blocks were then completed using the updated frequency discrimination threshold, and with the presentation of neurofeedback based on the output of the classifier on deviant trials. A final frequency discrimination threshold measurement was made thereafter.

Following the completion of these measurements, two or more additional blocks of EEG measurements were made using a 500 Hz standard stimulus and a 600 Hz deviant stimulus, which was expected to elicit large MMN responses given the salience of the stimulus contrasts. These data serve as the basis of the analyses performed. Two participants (2 and 14) completed three blocks of these measurements, while one participant (1) completed four blocks of measurements.

### Stimuli

The stimuli used in the experiment were pure sinusoidal tones with a duration of 100 ms and cosine on/offset envelopes of 5 ms. The fundamental frequency ($f0$) of the tones was always between 500-600 Hz.

### Sequences

EEG measurements were made using oddball sequences containing 85% standard stimuli (always 500 Hz) and 15% deviant stimuli (either threshold or 600 Hz) with a stimulus onset asynchrony (SOA) of 500 ms. Each block included a total of 150 deviant trials. These blocks lasted approximately 8.3 minutes. An additional block was measured prior to the classifier training that included 450 repetitions of the threshold stimulus.

## Dataset 2: Participant information and design

### Participants

12 participants (aged 18-38) reporting normal hearing and with normal or corrected-to-normal vision participated in the experiment. The participants were recruited through either through the Department of Cognitive Artificial Intelligence at the Donders Centre for Cognition (participants 1-6) or through the Radboud University Research participant system (participants 7-12). All participants provided informed consent prior to the experiment. The research was conducted with the approval of the ethics committee of the Faculty of Social Sciences at the Radboud University Nijmegen.

### Procedure

Participants completed four sessions of measurements on separate days within a week. At the beginning of the first and the end of the fourth day, participants completed three measurements of their frequency discrimination threshold, and three measurements of their

duration discrimination threshold, using the same type of staircase procedure as for Dataset 1. Participants also completed multiple blocks of EEG measurements on each day.

For the first six participants (1-6), two sets of measurement blocks were completed in the first three days: offline, and neurofeedback. Only neurofeedback measurements were made on the fourth day. In the offline measurements, five EEG blocks were measured and used for training a classifier for use in the neurofeedback portion of the experiment. At least five additional EEG blocks were measured in the neurofeedback portion of each session. For the second six participants (7-12), only neurofeedback measurements were made. Neurofeedback was generated using a classifier trained on the first six participant's data. A total of eight EEG blocks were measured in each of the four sessions.

For the analyses, only data from the first three days of measurements was utilized. The first three blocks of data recorded in each session were used in the analyses.

### Stimuli

The stimuli used in the experiment were harmonic tones containing two partials at $2 * f0$ (-3 dB) and $3 * f0$ (-6 dB) with a duration of 75 ms and cosine on/offset envelopes of 5 ms. The fundamental frequency ($f0$) of the tones was always 500 Hz, with the exception of the frequency deviant. Five types of deviant stimuli were used in a design based on Näätänen et al. (2004): frequency (550 Hz), amplitude ($\pm$10 dB), duration (25 ms), location ($\pm 800$ $\mu S$ ITD) and gap (25 ms ramped silence inserted between 25-50 ms).

### Sequences

Stimuli in the EEG measurement blocks were presented using optimal MMN sequences (Näätänen et al., 2004) containing 50% standard stimuli (500 Hz, 75 ms duration) and 10% of each type of deviant. Standard and deviant stimuli constantly alternated with an SOA of 500 ms, with all five types of deviants occurring within the span of 5 deviant trials in pseudorandom order. A total of 300 standard trials and 300 deviant trials were measured in each block. Each block lasted approximately 5 minutes.

## Common data acquisition and analysis methods

### EEG Data Acquisition

EEG was measured using a 64-channel BioSemi ActiveTwo amplifier and active AgCl electrodes at a sample rate of 2048 Hz. Electrode placement was performed according to the international 10-20 system. Additionally, horizontal/vertical EOG and left/right mastoid leads were recorded. Data acquisition was performed inside an electrically shielded and acoustically attenuated cabin.

### Stimulus Presentation

Auditory stimuli were presented via Etymotic ER-P4 insert headphones at approximately 70 dB SPL (excluding loudness deviants used during the collection of Dataset 2) inside an acoustically attenuated cabin using a Macintosh iMac computer via a MOTU 828 mk3 audio interface and the PsychToolbox MATLAB toolkit. Participants viewed silent films

on a 15.4 TFT display approximately 60-70 cm from their eyes during the measurements. Films were presented in a 300x400 pixel region centered within the 800x600 pixel display, in order to reduce eye movements to the corners of the screen.

## Data Pre-processing

The raw EEG data was pre-processed using the Fieldtrip software toolkit for MATLAB (Oostenveld et al., 2011). Data from all deviant trials and from the standard trials preceding them were selected for analysis. For Dataset 1, this was 30% of all recorded trials. All available trials were used for Dataset 2. Data were initially sliced and stored in epochs ranging from -200 to 600 ms relative to stimulus onset, and resampled to 256 Hz.

A bad-channel detection and repair procedure was then carried out on a per epoch basis. Channels with offsets exceeding $\pm 35\ mV$ or with 50 Hz power exceeding $1000\ \mu V^2$ were repaired using a spherical spline interpolation of the neighboring electrodes (Perrin et al., 1989).

Data were then baseline-corrected to the mean of each channel in the 50 ms period prior to stimulus onset, and an independent component analysis (ICA) using all available data from a given recording session was performed. The ICA was carried out using the infomax algorithm, as implanted by the 'runica' function of the EEGLab toolkit (Delorme and Makeig, 2004). The purpose of the ICA was to remove components from the data containing muscular artifacts such as eye movements (Jung et al., 2000), thus preserving a larger number of trials in the data that would otherwise be lost during the subsequent artifact rejection step. This has been shown to be a useful step when analyzing individual MMN data (Bishop and Hardiman, 2010). Only components accounting for greater than 1% of the total variance in the data were considered for removal. The mean variance across epochs was calculated for each component, and the overall mean across components was set as a threshold, with components above the threshold selected for removal. The topography of each component was inspected, and incremental adjustments to the threshold were made on a per participant basis to ensure that non-artifactual components (i.e. those containing auditory responses) were not removed from the data. Following the removal of the selected components, data were reprojected back onto the original EEG channels.

Following the ICA, data were band-pass filtered between 1 and 25 Hz. Epochs with activity in EEG channels exceeding $\pm 75\ \mu V$ were removed from the dataset. Data were then re-referenced to the average of the left and right mastoid leads.

# Appendix B

# Summary of neurofeedback pilot experiment with active task instruction

*based on Brandmeyer, A., Sadakata, M., Spyrou, L., McQueen, J.M. & Desain, P., 2013. Perceptual learning via decoded-EEG neurofeedback. Proceedings of the 5th International Brain Computer Interface Meeting, Asilomar, CA, USA.*

## Introduction

BCIs are most commonly developed for use as communication devices for patients with motor impairments (van Gerven et al., 2009). Recently, it has been shown that the same multivariate methods underlying many BCIs can be used to provide novel forms of fMRI-based neurofeedback (NFB) that induce visual perceptual learning in healthy users (Shibata et al., 2011). An experiment was conducted to determine whether a similar approach is possible using NFB based on decoded auditory evoked potentials (AEPs) measured using EEG and an active NFB task instruction. Stimulus sequences containing high-probability 'standard' trials and low-probability 'deviant' trials were used. Deviant trials in these sequences are known to elicit both enhanced P3a responses and the mismatch negativity (MMN). It was hypothesized that providing NFB on the differing pattern of AEPs to standards and deviants would lead to a relative enhancement of these components, along with an induced perceptual learning effect that would improve behavioral discrimination.

## Material and Methods

### Participants

Six participants completed four days of testing. All participants reported normal hearing.

### Stimuli

Harmonic sinusoidal tones were used for the behavioral, offline EEG measurement and NFB portions of the experiment. During both the offline and NFB portions, so-called 'optimal' MMN sequences containing 5 types of deviant stimuli were utilized, as described in Näätänen et al. (2004). Two of these deviant stimuli served as NFB targets: duration (25 ms vs. 75 ms standard) and frequency (550 Hz vs. 500 Hz standard). Half of the participants' NFB was based on single-trial AEPs measured for duration deviants and the standard stimuli immediately preceding them, while the other half received NFB on frequency deviants and the standard stimuli preceding them.

### Procedure

At the beginning of the first day, participants completed measurements of their behavioral discrimination thresholds for both frequency and duration using a 2AFC staircase procedure. For the first three days, participants completed offline EEG measurements while watching silent films, followed by a NFB session in which they were instructed to make active use of the neurofeedback signal. On the fourth day, participants completed a NFB session followed by the same behavioral measurements as on the first day.

### Classification analysis and NFB parameters

On the first three days, individual data collected during the initial EEG measurements were used to train two quadratically regularized linear logistic regression classifiers. The fourth day made use of classifiers trained on the previous three days' worth of individual data. A first classifier was trained on a binary problem consisting all standard trials vs. all deviant trials, and was applied during NFB sessions to the standard trials preceding the target deviants. A second binary classifier was trained on trials for the NFB target (frequency or duration deviant) vs. an equal number of trials in which the same stimulus was measured in an isochronous sequence. During NFB, this classifier was applied to the target deviant trials. The outputs of the two classifiers for the previous five trials were combined into a single value during the NFB sessions to control the amount of blurriness of the films viewed by participants during the NFB sessions.

## Results

### Behavioral

Behavioral results can be seen in Figure B.1a. Thresholds for frequency discrimination in the post-test were significantly lower than in the pre-test ($p < .01$), with 5 out of 6 participants showing reduced thresholds in the post-test. No effects were observed for duration thresholds, nor were any training group effects observed.

### ERP measurements

ERP plots across the 4 days can be seen in Figure B.1b. For each day and measurement (offline vs. NFB), the mean amplitude of the MMN and P3a components in the individual

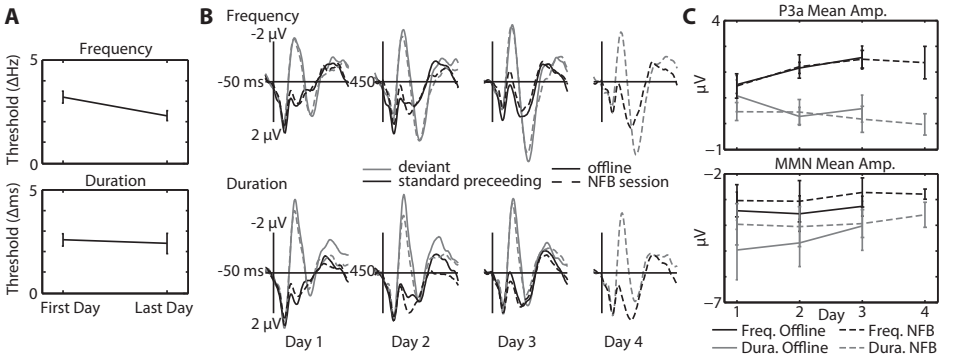


Figure B.1: *Behavioral and ERP results. A) Frequency and duration discrimination thresholds on the first and last days for all participants. B) Grand-average ERPs at fronto-central electrode locations across days and deviant types. C) Mean amplitudes of MMN and P3a components across days, session and deviant types*

grand average deviant  standard difference waveforms were computed in a 50 ms window around their peaks using the mean of nine fronto-central electrode locations where AEPs tend to be maximal. These results are plotted in Figure B.1c. For the P3a component, a significant effect of deviant type ($p < .05$) and an interaction of deviant type with measurement day ($p < .001$) were found, with higher P3a amplitudes for 5 out of 6 participants on the final two days (relative to the first day). No significant effects were found for the MMN component.

# Discussion

Regardless of which deviant was used to provide NFB to participants, a general enhancement of P3a responses to frequency deviants was observed along with an improvement in frequency discrimination performance. One possible explanation is that effortful control of the NFB mechanism leads to top-down modulation of perceptual processes related to the perception of deviant stimuli, and that this modulation is most apparent in responses to frequency deviants. While additional research is needed to clarify these effects, the present results suggest the potential for novel BCI applications for auditory perceptual learning in healthy users.

# Nederlandse Samenvatting

We leven in een wereld van geluid: de stemmen van onze familie en vrienden, de noten van onze favoriete muziek of de bekende geluiden van onze omgeving tijdens de dagelijkse routines. Deze geluiden worden zonder moeite of gedachte herkend, maar eigenlijk ligt er een zeer ingewikkeld neurofysiologisch systeem achter deze auditieve ervaringen. Gedurende ons leven leren wij de geluiden in onze directe omgeving te herkennen, zoals de fonemen van onze moedertaal of de tonen in de stemming van onze muziekcultuur.

Dit leerproces gaat samen met veranderingen in de synaptische verbindingen tussen neuronen. De effecten van leren kunnen gezien worden in als verschillen in de patronen van activiteit die worden gemeten door technieken zoals elektro-encefalografie (EEG) of functionele kernspintomografie (fMRI). Bij de geboorte heeft men nog geen ervaring met een specifieke taal, maar tussen zes en twaalf maanden is het al te merken dat babys een voorkeur voor fonemen in hun moedertaal hebben. Als volwassene heeft men de fonetische categorieën van zijn of haar moedertaal geleerd, en in EEG kan dit gezien worden als we middelen over herhaalde metingen en de zogenoemde event-related potential (ERP) bekijken.

De mismatch negativity, of kortweg MMN, is een ERP-component dat wordt gemeten door middel van zogenaamde oddball-reeksen geluiden. Hierbij zijn twee typen klanken te horen: een standaard stimulus (ongeveer 85% van de geluiden) en, op onregelmatige momenten, een afwijkende (deviant) stimulus. Deze twee geluiden kunnen bijvoorbeeld twee verschillende fonemen zijn, of twee muzikale tonen met verschillende toonhoogtes. De gemeten ERPs van deze twee geluiden hebben een karakteristiek verschil tussen 100-300 milliseconden, waarbij de ERP gemeten tijdens de deviant een grotere negatieve piek laat zien. Als de ERP voor het standaard geluid van de ERP voor de deviant afgetrokken wordt, kan de MMN-component gezien worden.

Een van de interessantste aspecten van de MMN-component is dat het verschillen in de manier waarop de hersenen van specifieke groepen mensen geluid verwerken laat zien. Dit wordt bevestigd door de resultaten van veel experimenten waarin mensen luisteren naar fonetische contrasten in een vreemde taal. Als de twee geluiden binnen een fonetische categorie in de moedertaal van deze mensen vallen, dan is de MMN-component verminderd of helemaal niet te zien. Deze verschillen in hersenactiviteit gaan samen met verschillen in gedrag van deze groepen bij het laten horen en reageren op diezelfde fonemen. Dit patroon geldt ook voor het waarnemen van muzikale geluiden door muzikanten en mensen zonder muzikale achtergrond. Dus blijkt de MMN de neurofysiologische basis van belangrijke auditieve perceptuele processen te kunnen weergeven.

De MMN-component wordt bestudeerd door het meten van ERPs, een gemiddelde van de gemeten hersenactiviteit van meerdere (50-300) herhalingen van een bepaalde prikkel

of gebeurtenis op een vast tijdstip. In de afgelopen jaren zijn methoden ontwikkeld in het onderzoeksveld van kunstmatige intelligentie om bepaalde EEG componenten, zoals de P300-respons, in real-time of online te analyseren. Dit soort analyses, die patroonclassificatie heten, vormen de basis van hersen-computer interfaces (BCIs). Zulke BCIs bieden de mogelijkheid om een apparaat (bijvoorbeeld een communicatiesysteem) aan te sturen met vormen van mentale activiteit. Dit kan een belangrijke oplossing zijn voor patiënten met spierziektes zoals amyotrofe laterale sclerose (ALS), omdat, tijdens het verloop van deze ziektes, men de capaciteit verliest om de spieren te bewegen, en daarmee de mogelijkheid om te communiceren.

Terwijl de ontwikkeling van BCIs in het algemeen gericht is op specifieke groepen patiënten, zijn er ook toepassingen voor BCI-technieken met gezonde gebruikers. Een BCI-systeem gebaseerd op het meten van MMN-responsen zou individuele verschillen in het waarnemen van taal of muziek kunnen detecteren in real-time, in het kader van onderwijs of voor klinische toepassingen. Ook zou dit soort analyse van MMN-responsen tot een nieuwe vorm van neurofeedback kunnen leiden. Neurofeedback is een methode waarin metingen van de hersenactiviteit van een gebruiker direct gepresenteerd worden als visueel of auditief signaal, met de bedoeling dat de gebruiker bewust probeert om dit signaal te beheersen. Uit recent onderzoek blijkt dat het mogelijk is om visuele waarnemingen te verbeteren door het gebruik van fMRI-neurofeedback. Een logische vraag is of een vergelijkbare vorm van neurofeedback gebaseerd op de MMN-component een positieve invloed op auditief perceptueel leren heeft.

Dit proefschrift laat de resultaten zien van onderzoek gericht op het ontwikkelen van een neurofeedbackparadigma voor auditief perceptueel leren. Vier empirische hoofdstukken stellen de volgende onderzoeksvragen: wat zijn de verschillen tussen eerste- en ervaren tweede-taalsprekers qua gedrag en hersenactiviteit in het categorische auditieve waarnemen van fonemen (Hoofdstuk 2)? Is het mogelijk om de verschillen in hersenactiviteitpatronen (bijvoorbeeld de MMN-respons) tussen deze groepen taalsprekers te analyseren op het niveau van individuele metingen, zoals in het BCI-veld wordt gedaan (Hoofdstuk 3)? Wat zijn de gevolgen van neurofeedback op de MMN-respons voor auditieve waarnemingen en hersenactiviteit (Hoofdstuk 4)? Hoe kan de patroonclassificatie-analyse van EEG-metingen van de MMN-respons geoptimaliseerd worden in de context van bepaalde toepassingen, zoals neurofeedback en BCI, en wat voor resultaten kan men verwachten in dit soort situaties (Hoofdstuk 5)? In Hoofdstuk 6 worden bovengenoemde resultaten samengevat, en de uitdagingen en openstaande vragen voor dit onderzoek bediscussieerd.

Hoofdstuk 2 bevat de resultaten van een experiment waarmee werd onderzocht wat de verschillen zijn tussen twee groepen in het waarnemen van Engelstalige fonemen: mensen met Engels als moedertaal (11 proefpersonen), en mensen met Nederlands als moedertaal die ook ervaring met Engels als tweede taal hebben (11 proefpersonen met tenminste zes jaren Engelse les op school). Dit experiment gebruikte een continuüm van zeven spraakgeluiden van "pa" tot "ba", omdat er een interessant verschil bestaat tussen de zogenoemde "plosieve" medeklinkers in het Engels en in het Nederlands: terwijl de Nederlandse /b/ en /p/ van elkaar worden onderscheiden door een pre-vocalisatie in het /b/-geluid voor het plosief portie van het medeklinker, verschillen de Engelse /b/ en /p/ in het gebruik van aspiratie (een soort aanblazing, zoals een zacht /h/ geluid) na de plosief portie van de stemloze /p/. Het verschil in aspiratie tussen de /p/ en /b/ fonemen leidt naar een verschil in de lengte van deze geluiden (voice-onset time, oftwel VOT). Het Nederlands maakt geen gebruik van aspiratie.

De vraag is of de Nederlandse proefpersonen in dit experiment hebben geleerd om de verschillen in aspiratie in deze Engelstalige geluiden waar te nemen. Het experiment maakte gebruik van gedragsmetingen in de vorm van een categorisatietaak (is het geluid een /pa/ of een /ba/?) en van EEG-metingen van de MMN-respons. De resultaten van dit experiment lieten zien dat de Nederlandse proefpersonen minder gevoelig waren voor de hoeveelheid aspiratie in de spraakgeluiden dan de Engelstalige proefpersonen. Dit verschil tussen de twee groepen trad op in gedragsmetingen: een van de geluiden werd duidelijk als /ba/ gehoord door de Engelstalige proefpersonen, maar door de Nederlandstalige proefpersonen werd deze geluid nog steeds de helft van de tijd als /pa/ gehoord. De Engelstalige proefpersonen toonden ook grotere MMN-responsen dan de Nederlandse proefpersonen voor een binnen-categorie geluidscontrast en voor een contrast dat op de categorie grens van hun eigen moedertaal. Deze resultaten laten zien dat, in het geval dat de akoestische aanwijzingen voor spraakgeluiden in een eerste en tweede taal niet overeenkomen, de ervaring die we in onze vroege leven krijgen met onze moedertaal een belangrijk rol blijft spelen.

De EEG-data van dit experiment vormen de basis voor een uitgebreide patroonclassificatie-analyse die in Hoofdstuk 3 wordt beschreven. Hier wordt gekeken of de individuele en groepsniveauverschillen in het categorische waarnemen van de Engelstalige fonemen te zien zijn in de single-trial EEG-metingen. Dit is een centrale vraag voor dit proefschrift, omdat de voorgestelde BCI- en neurofeedbacksystemen gevoelig moeten zijn voor verschillen in een of enkele MMN-responsen om effectief te zijn. Uit de resultaten blijkt dat de uitkomsten van de classificatieanalyse de individuele verschillen in gedrag reflecteren, en dat de hersenactiviteit die wordt gemeten met een binnen-categoriecontrast anders is voor de twee taalgroepen: uitkomsten waren significant hoger voor de Engelstalige proefpersonen dan voor de Nederlandse proefpersonen. Dit werd ook bevestigd door de resultaten van een classificatieanalyse tussen proefpersonen. Tenslotte blijkt het ook mogelijk te zijn om de uitkomsten van de classificatieanalyse te verbeteren door het gebruik van enkele EEG-metingen in plaats van eén. Deze resultaten tonen aan dat een BCI-systeem gebaseerd op MMN-responsen, bijvoorbeeld voor neurofeedbacktoepassingen, realistisch is, en laten specifieke parameters zien die gebruikt kunnen worden in een online paradigma.

Hoofdstuk 4 beschrijft een experiment waarin de effecten van een vorm van neurofeedback op auditieve perceptuele gevoeligheid en hersenactiviteit ondergezocht worden. Deze neurofeedback is gebaseerd op EEG-metingen in een MMN-paradigma met eenvoudige tonen op verschillende toonhoogtes. Dezelfde patroonclassificatiemethode van Hoofdstuk 3 werd gebruikt om MMN-responsen op individuele geluiden te detecteren. De uitkomst van deze classificaties stuurde een visuele vorm van neurofeedback (de beeldscherpte van een film) aan. Een groep van 8 proefpersonen onderging vier dagen van neurofeedbacktrainingen, samen met gedragsmetingen van hun perceptuele gevoeligheid voor kleine verschillen in toonhoogte. Een tweede controlegroep onderging dezelfde procedure, maar kreeg willekeurige feedback ongerelateerd aan hun hersenactiviteit. Hoewel er geen effecten van de neurofeedback op perceptuele gevoeligheid waren gevonden (beide groepen toonden hetzelfde patroon van verbetering tussen dag 1 en dag 4), werd er wel een effect van de neurofeedback op de amplitude van de MMN-respons tijdens de neurofeedbacksessies gevonden: MMN-responsen waren groter voor de feedbackgroep dan voor de controlegroep. Dit geeft aan dat deze vorm van decoded-EEG neurofeedback een effect kan hebben op hersenactiviteit tijdens auditieve waarnemingen.

Een algemene methode voor de toepassing van patroonclassificatie-analyses in de con-

text van het MMN-paradigma, neurofeedback en BCI wordt beschreven in Hoofdstuk 5. Ten eerste wordt de verzamelde data van twee experimenten gebruikt om de parameters van EEG-data pre-processen te optimaliseren voor classificatie. Deze analyses wijzen naar specifieke parameters, zoals de hoge en lage filterfrequenties en digitale samplingfrequentie, die de resultaten van de classificatieanalyse kunnen verbeteren. Daarna worden drie verschillende methoden (Searchlight procedure, AUC curves, en ERP Component-based selection) gebruikt om de invloed van verschillende aspecten van de data op de resultaten van de classificatie-analyse duidelijk te maken. In het volgende gedeelte wordt de mogelijke generalisatie van de classificatieresultaten naar nieuwe data en individuen geanalyseerd. Deze resultaten worden ook geïnterpreteerd als een probabilistisch signaal dat doorlopende veranderingen in de individuele EEG-metingen kan detecteren. Tot slot wordt het gebruik van deze probabilistische interpretatie van de classificatie-analyse bediscussieerd in het kader van de beoogde gebruiksscenarios: BCI en neurofeedback.

De bovengenoemde resultaten zijn in Hoofdstuk 6 kort samengevat, samen met een discussie van specifieke uitdagingen voor deze vorm van neurofeedback en opvolgende onderzoeksvragen. In het bijzonder wordt over de rol van de classificatie-uitkomsten en de specifieke geluidsprikkels in de effectiviteit van de neurofeedback besproken. Deze twee aspecten van het neurofeedbackparadigma zijn de uitgangspunten voor nieuwe onderzoeksvragen: 1) Kan het gebruik van grotere datasets met meerdere proefpersonen voor de classificatie-analyse leiden naar betere uitkomsten en de ondersteuning van verschillende geluidsprikkels tijdens neurofeedback, zoals fonemen? 2) Wat is de relatie tussen het perceptuele leerproces in taal- en muziekonderwijs en de effecten van de neurofeedback op hersenactiviteit?

Vanuit een breder perspectief worden de ethische aspecten van neurofeedback en BCI kort besproken, in de context van de potentiele voor- en nadelen van dit soort technologieën voor de maatschappij. Aan de ene kant bieden deze technologieën de mogelijkheid om bepaalde groepen mensen, zoals patiënten en studenten, te ondersteunen. Aan de andere kant kunnen zulke systemen gebruikt worden om individuen te stigmatiseren of grenzen van individuele privacy te vervagen. Dus moeten deze kwesties goed in het oog gehouden worden door onderzoekers en ontwikkelaars van deze technologie, maar ook door het brede publiek.

Al met al heeft dit proefschrift laten zien hoe wij een beter inzicht kunnen krijgen in de neurofysiologische basis van auditief perceptueel leren door het gebruik van patroonclassificatie-analyse en real-time-paradigmas. Methoden die een steeds belangrijker rol kunnen spelen in het onderzoek en onderwijs van de toekomst.

# Bibliography


Alain, C., Snyder, J. S., He, Y., and Reinke, K. S. (2006). Changes in auditory cortex parallel rapid perceptual learning. *Cerebral Cortex*, 17(5):1074–1084.

Andersson, P., Pluim, J. P. W., Siero, J. C. W., Klein, S., Viergever, M. A., and Ramsey, N. F. (2011). Real-Time decoding of brain responses to visuospatial attention using 7T fMRI. *PLOS ONE*, 6(11):e27638.

Asaridou, S. S. and McQueen, J. M. (2012). Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. *Frontiers in Psychology*, 4(321):1–14.

Atienza, M. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning & Memory*, 9(3):138–150.

Badgaiyan, R. D. (2013). Detection of dopamine neurotransmission in "real time". *Frontiers in Neuroscience*, pages 1–4.

Bastiaansen, M., Mazaheri, A., and Jensen, O. (2011). Beyond ERPs: Oscillatory neuronal dynamics. In Luck, S. J. and Kappenman, E. S., editors, *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press.

Belitski, A., Farquhar, J., and Desain, P. (2011). P300 audio-visual speller. *Journal of Neural Engineering*, 8(2):025022.

Best, C. (1993). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In Boysson-Bardies, B. D., de Schonen, S., Jusczyk, P., and MacNeilage, P., editors, *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, pages 167–224. Kluwer, Dordrecht, Netherlands.

Best, C., McRoberts, G. W., and Sithole, N. M. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14:345–360.

Bishop, C. (2009). *Pattern Recognition and Machine Learning*. Springer New York.

Bishop, D. and Hardiman, M. (2010). Measurement of mismatch negativity in individuals: a study using single-trial analysis. *Psychophysiology*, 47(4):1–9.

Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychological Bulletin*, 133(4):651–672.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components - a tutorial. *NeuroImage*, 56(2):814–825.

Blankertz, B., Sannelli, C., Halder, S., Hammer, E. M., Kübler, A., Müller, K. R., Curio, G., and Dickhaus, T. (2010). Neurophysiological predictor of SMR-based BCI performance. *NeuroImage*, 51(4):1303–1309.

Brandmeyer, A., Desain, P., and McQueen, J. (2012). Effects of native language on perceptual sensitivity to phonetic cues. *NeuroReport*, 23:653–657.

Brandmeyer, A., Farquhar, J., McQueen, J., and Desain, P. (2013). Decoding speech perception by native and non-native dpeakers using single-trial electrophysiological data. *PLOS ONE*, 8(7):e68261.

Brattico, E., Pallesen, K. J., Varyagina, O., Bailey, C., Anourova, I., Järvenpää, M., Eerola, T., and Tervaniemi, M. (2009). Neural discrimination of nonprototypical chords in music experts and laymen: an MEG study. *Journal Of Cognitive Neuroscience*, 21(11):2230–2244.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA.

Brouwer, A.-M., Hogervorst, M. A., van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering*, 9(4):045008.

Brunner, C., Naeem, M., Leeb, R., Graimann, B., and Pfurtscheller, G. (2007). Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. *Pattern Recognition Letters*, 28:957–964.

Bukard, R. F., Don, M., and Eggermont, J. J., editors (2007). *Auditory Evoked Potentials: Basic Principles and Clinical Applications*. Philadelphia, PA: Wolters Kluwer Health.

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Cardinal, R. (2006). Neural systems implicated in delayed and probabilistic reinforcement. *Neural Networks*, 19(8):1277–1301.

Chen, Y., Namburi, P., Elliott, L. T., Heinzle, J., Soon, C. S., Chee, M. W., and Haynes, J.-D. (2011). Cortical surface-based searchlight decoding. *NeuroImage*, 56(2):582–592.

Collins, B. and Mees, I. (1996). *The Phonetics of English and Dutch, 3rd edition*. Brill.

Conway, C. M., Bauernschmidt, A., Huang, S. S., and Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3):356–371.

Cormack, G. V. (2007). Email spam filtering: a systematic review. *Foundations and Trends in Information Retrieval*, 1(4):335–455.

Coyle, D., Principe, J., Lotte, F., and Nijholt, A. (2013). Guest Editorial: Brain/neuronal - computer game interfaces and interaction. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):77–81.

Coyle, S. M., Ward, T. E., and Markham, C. M. (2007). Brain–computer interface using a simplified functional near-infrared spectroscopy system. *Journal of Neural Engineering*, 4(3):219–226.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, 116:3668–3678.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D. G., and Acharyya, M. (2005). Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *NeuroImage*, 28(3):663–668.

de Charms, R. C. (2008). Applications of real-time fMRI. *Nature Reviews*, 9(9):720–729.

de Vries, M. H., Barth, A. C., Maiworm, S., Knecht, S., Zwitserlood, P., and Flöel, A. (2010). Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar. *Journal Of Cognitive Neuroscience*, 22(11):2427–2436.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, 8(4):919–924.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Desain, P. and Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32(3):341–365.

Duda, R., Hart, P., and Stork, D. (2001). *Pattern Classification*. New York, NY: John Wiley and Sons.

Duncan, C., Barry, R., Connolly, J., Fischer, C., Michie, P., Naatanen, R., Polich, J., Reinvang, I., and van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, 120(11):1883–1908.

Farquhar, J. R. D. and Hill, J. (2013). Interactions between pre-processing and classification methods for event-related-potential classification. *Neuroinformatics*, 11(2):175–192.

Farwell, L. A. and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70:510–523.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874.

Fischer, C., Luauté, J., Adeleine, P., and Morlet, D. (2004). Predictive value of sensory and cognitive evoked potentials for awakening from coma. *Neurology*, 63(4):669–673.

Formisano, E., De Martino, F., Bonte, M., and Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, 322(5903):970–973.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2):78–84.

Fruhstorfer, H. (1971). Habituation and dishabituation of the human vertex response. *Electroencephalography and Clinical Neurophysiology*, 30(4):306–312.

Fuentemilla, L. L., Marco-Pallares, J., Munte, T. F., , and Grau, C. (2008). Theta EEG oscillatory activity and auditory change detection. *Brain Research*, 1220:93–101.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal Of Cognitive Neuroscience*, 16(6):1010–1021.

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology*, 120(3):453–463.

Gonsalvez, C. and Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology*, 39(3):388–396.

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., and Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, 76(11):1006–1014.

Goswami, U. (2006). Neuroscience and education: from research to practice? *Nature Reviews Neuroscience*, 7(5):406–413.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. *Neuropsychologia*, 9:317–323.

Green, D. and Swets, J. (1974). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Griffiths, T. D. and Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5(11):887–892.

Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., and Rao, R. P. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 835–844.

Grimm, S., Escera, C., Slabu, L., and Costa-Faidella, J. (2011). Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. *Psychophysiology*, 48(3):377–384.

Halder, S., Rea, M., Andreoni, R., Nijboer, F., Hammer, E. M., Kleih, S. C., Birbaumer, N., and Kübler, A. (2010). An auditory oddball brain–computer interface for binary choices. *Clinical Neurophysiology*, 121(4):516–523.

Hammond, D. (2011). What is neurofeedback: An update. *Journal of Neurotherapy*, 15(4):305–336.

Handy, T. C. (2005). *Event-Related Potentials: A Methods Handbook*. Cambridge, MA: MIT Press.

Haselager, P., Vlek, R., Hill, J., and Nijboer, F. (2009). A note on ethical aspects of BCI. *Neural Networks*, pages 1–6.

Haynes, J.-D. and Rees, G. (2005a). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, 8(5):686–691.

Haynes, J.-D. and Rees, G. (2005b). Predicting the stream of consciousness from activity in human visual cortex. *Current Biology*, 15(14):1301–1307.

Haynes, J.-D. and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7):523–534.

Herrmann, B., Maess, B., Kalberlah, C., Haynes, J., and Friederici, A. (2012). Auditory perception and syntactic cognition: Brain activity-based decoding within and across subjects. *European Journal of Neuroscience*, 35(9):1488–1496.

Hill, N. J., Lal, T., Schröder, M., Hinterberger, T., Widman, G., Elger, C., Schölkopf, B., and Birbaumer, N. (2006). *Classifying event-related desynchronization in EEG, ECoG and MEG signals*, pages 404–413. Springer Berlin Heidelberg.

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., Haddadin, S., Liu, J., Cash, S. S., van der Smagt, P., and Donoghue, J. P. (2013). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375.

Höhne, J. (2011). A novel 9-class auditory ERP paradigm driving a predictive text entry system. *Frontiers in Neuroscience*, 5:1–10.

Höhne, J., Krenzlin, K., Dähne, S., and Tangermann, M. (2012). Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *Journal of Neural Engineering*, 9(4):045003.

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87:B47–B57.

Jensen, O. and Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *European Journal of Neuroscience*, 15(8):1395–1399.

Johnston, S. J., Boehm, S. G., Healy, D., Goebel, R., and Linden, D. E. J. (2010). Neurofeedback: A promising tool for the self-regulation of emotion networks. *NeuroImage*, 49(1):1066–1072.

Jung, T., Makeig, S., Humphries, C., Lee, T., McKeown, M., Iragui, V., and Sejnowski, T. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2):163–178.

Kamitani, Y. and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8(5):679–685.

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., Itoh, K., Koshida, I., Yumoto, M., Iramina, K., Kato, N., and Ueno, S. (2001). Brain lateralization for mismatch response to across- and within-category change of vowels. *NeuroReport*, 12(11):2467–2471.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., and Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116(3):321–340.

Kay, K. N., Naselaris, T., Prenger, R. J., and Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452:352–355.

Keating, P., Mikos, M., and Ganong, W. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, 70(5):1262–1271.

Keenan, J. P., Thangaraj, V., Halpern, A. R., and Schlaug, G. (2001). Absolute Pitch and Planum Temporale. *NeuroImage*.

Kindermans, P. and Schrauwen, B. (2013). Dynamic stopping in a calibration-less P300 speller . In *Proceedings of the 5th International Brain-Computer Interface Meeting*.

Kleih, S. C., Nijboer, F., Halder, S., and Kübler, A. (2010). Motivation modulates the P300 amplitude during brain-computer interface use. *Clinical Neurophysiology*, 121(7):1023–1031.

Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport*, 10(6):1309–1313.

Koelsch, S. and Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Science*, 9(12):578–584.

Korpilahti, P., Krause, C. M., Holopainen, I., and Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, 76(3):332–339.

Korpilahti, P., Lang, H., and Aaltonen, O. (1995). Is there a late-latency mismatch negativity (MMN) component? *Electroencephalography and Clinical Neurophysiology*, 95(4):96P.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, 126(1):159.

Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5:831–843.

Kuhl, P., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2):F13–F21.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 31:606–608.

Kujala, T., Tervaniemi, M., and Schroeger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, 74(1):1–19.

Lansbergen, M. M., van Dongen-Boomsma, M., Buitelaar, J. K., and Slaats-Willemse, D. (2011). ADHD and EEG-neurofeedback: a double-blind randomized placebo-controlled feasibility study. *Journal of Neural Transmission*, 118(2):275–284.

Leeb, R., Friedman, D., Müller-Putz, G. R., Scherer, R., Slater, M., and Pfurtscheller, G. (2007). Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. *Computational Intelligence and Neuroscience*, 79642:1–8.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49:467.

Li, Y., Long, J., Yu, T., Yu, Z., Wang, C., Zhang, H., and Guan, C. (2010). An EEG-based BCI system for 2-D cursor control by combining mu/beta Rhythm and P300 potential. *Biomedical Engineering, IEEE Transactions on*, 57(10):2495–2505.

Liberman, A., Harris, K., Kinney, J., and Lane, H. (1961). The discrimination of relative-onset time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology*, 61:379–388.

Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J., and Stephan, K. E. (2013). Modelling trial-by-trial changes in the mismatch negativity. *PLOS Computational Biology*, 9(2):e1002911.

Lisman, J. E. and Jensen, O. (2013). The theta-gamma neural code. *Neuron*, 77(6):1002–1016.

Luck, S. J. and Kappenman, E. S. (2011). *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press.

Lynch, M. P. and Eilers, R. E. (1991). Children's perception of native and non-native musical scales. *Music Perception: An Interdisciplinary Journal*, 9(1):121–131.

Lynch, M. P., Eilers, R. E., Kimbrough Oller, D., and Urbano, R. C. (1990). Innateness, experience and music perception. *Psychological Science*, 1(4):272–276.

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., and Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, 16(1):68–90.

Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8(5):204–210.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190.

Martin, B. A., Tremblay, K. L., and Stapells, D. R. (2007). Principles and applications of cortical auditory evoked potentials. In Bukard, R. F., Don, M., and Eggermont, J. J., editors, *Auditory Evoked Potentials: Basic Principles and Clinical Applications*, chapter 23, pages 482–507. Wolters Kluwer Health.

May, P. C. and Tiitinen, H. (2009). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, 46:1–57.

McFarland, D. J. and Wolpaw, J. R. (2005). Sensorimotor rhythm-based brain-computer interface (BCI): feature selection by regression improves performance. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(3):372–379.

Mees, I. and Collins, B. (1982). A phonetic description of the consonant system of Standard Dutch (ABN). *Journal of the International Phonetics Association*, 12:2–12.

Meltzoff, A. N., Kuhl, P. K., Movellan, J., and Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, 325(5938):284–288.

Menning, H., Roberts, L., and Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport*, 11(4):817–822.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219:36–47.

Michie, P. T. (2001). What has MMN revealed about the auditory system in schizophrenia? *International Journal of Psychophysiology*, 42(2):177–194.

Miltenberger, R. G. (2011). *Behavior Modification: Principles and Procedures: Principles and Procedures*. Wadsworth, Cengage Learning.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975). An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18:331–340.

Mori, S., Nishida, H., and Yamada, H. (1999). *Optical Character Recognition*. John Wiley and Sons.

Muller, K., Anderson, C. W., and Birch., G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):165–169.

Müller-Putz, G., Scherer, R., Brunner, C., R, L., and Pfurtscheller, G. (2008). Better than random? a closer look on BCI results. *International Journal of Bioelectromagnetism*, 10(1):52–55.

Munka, L. and Berti, S. (2006). Examining task-dependencies of different attentional processes as reflected in the P3a and reorienting negativity components of the human event-related brain potential. *Neuroscience Letters*, 396:177–181.

Münte, T. F., Alltenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews*, 3:473–478.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear and Hearing*, 16:6–18.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38(1):1–21.

Näätänen, R., Gaillard, A., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42(4):313–329.

Näätänen, R., Jacobsen, T., and Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, 42:25–32.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huitilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385:432–434.

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, 118(12):2544–2590.

Näätänen, R., Pakarinena, S., Rinnea, T., and Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology*, 115(1):140–144.

Nijboer, F., Sellers, E. W., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., Halder, S., Mochty, U., Krusienski, D. J., Vaughan, T. M., Wolpaw, J. R., Birbaumer, N., and Kübler, A. (2008). A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119(8):1909–1916.

Ogar, J. M., Baldo, J. V., Wilson, S. M., and Brambati, S. M. (2011). Semantic dementia and persisting Wernicke's aphasia: Linguistic and anatomical profiles. *Brain and Language*.

Oostenveld, R., Fries, P., Maris, E., , and Schoffelen, J. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011:1–9.

Owen, A. M. (2006). Detecting awareness in the vegetative state. *Science*, 313(5792):1402–1402.

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford University Press.

Patel, A. D. and Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, 11(9):369–372.

Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, 45(1):S199–S209.

Peretz, I. and Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6(7):688–691.

Perrin, F., Pernier, J., Bertrand, O., and Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, 72:184–187.

Pham, M., Hinterberger, T., Neumann, N., Kübler, A., Hofmayer, N., Grether, A., Wilhelm, B., Vatine, J., and Birbaumer, N. (2005). An auditory brain-computer interface based on the self-regulation of slow cortical potentials. *Neurorehabilitation and Neural Repair*, 19(3):206–218.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, 12(6):1038–1055.

Pisoni, D. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2):253–260.

Polich, J. (1989). Neuroanatomical contributions to individual differences in p300 morphology. In *Brain Dynamics*, pages 331–338. Springer Berlin Heidelberg.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128–2148.

Rammsayer, T. and Altenmüller, E. (2006). Temporal information processing in musicians and non-musicians. *Music Perception*, 24(1):37–48.

Ramsey, N. F., Heuvel, M. P. V. D., Kho, K. H., and Leijten, F. S. S. (2006). Towards human BCI applications based on cognitive brain systems: an investigation of neural signals recorded from the dorsolateral prefrontal cortex. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):214–217.

Reinke, K., He, Y., Wang, C., and Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research*, 17(3):781–791.

Ritter, W., Vaughan, H., and Costa, L. (1968). Orienting and habituation to auditory stimuli: a study of short terms changes in average evoked responses. *Electroencephalography and Clinical Neurophysiology*, 25(6):550–556.

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., and Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research*, 111(1-2):13–23.

Rüschemeyer, S.-A., Zysset, S., and Friederici, A. D. (2006). Native and non-native reading of sentences: an fMRI experiment. *NeuroImage*, 31(1):354–365.

Sadakata, M., Ohgushi, K., and Desain, P. (2004). A cross-cultural comparison study of the production of simple rhythmic patterns. *Psychology of Music*, 32(4):389–403.

Salvaris, M. and Sepulveda, F. (2009). Visual modifications on the P300 speller BCI paradigm. *Journal of Neural Engineering*, 6(4):046011.

Sams, M., Alho, K., and Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the ERP. *Psychophysiology*, 21(4):434–441.

Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Cinical Neurophysiology*, 62:437–448.

Schaefer, R., Farquhar, J., Blokland, Y., and Sadakata, M. (2011). Name that tune: Decoding music from the listening brain. *Neuroimage*, 56(2):843–849.

Schalk, G., McFarland, D., Hinterberger, T., Birbaumer, N., and Wolpaw, J. (2004). BCI2000: A general purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043.

Scharnowski, F., Hutton, C., Josephs, O., Weiskopf, N., and Rees, G. (2012). Improving visual perception through neurofeedback. *Journal of Neuroscience*, 32(49):17830–17841.

Schlaug, G., Jäncke, L., Huang, Y., and Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, 267(5198):699–701.

Schmithorst, V. J. and Holland, S. K. (2003). The effect of musical training on music processing: A functional magnetic resonance imaging study in humans. *Neuroscience Letters*, 348(2):65–68.

Schreuder, M., Blankertz, B., and Tangermann, M. (2010). A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue. *PLOS ONE*, 5(4):e9813.

Schroger, E. and Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *NeuroReport*, 9:3355–3358.

Sellers, E. and Donchin, E. (2006). A P300-based brain-computer interface: Initial tests by ALS patients. *Clinical Neurophysiology*, 117:538–548.

Sellers, E. W. (2013). New horizons in brain-computer interface research. *Clinical Neurophysiology*, 124(1):2–4.

Shanks, D. R. (2010). Learning: from association to cognition. *Annual Review of Psychology*, 61(1):273–301.

Sharma, A. and Dorman, M. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America*, 107(5):2697–2703.

Sharma, A. and Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, 106(2):1078–1083.

Shibata, K., Watanabe, T., Sasaki, Y., and Kawato, M. (2011). Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science*, 334(6061):1413–1415.

Siegel, J. A. and Siegel, W. (1977). Categorical perception of tonal intervais: Musicians can't tell sharp from flat. *Perception and Psychophysics*, 21(5):399–407.

Smith, M. E., Gevins, A., Brown, H., Karnik, A., and Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(3):366–380.

Stalinski, S. M. and Schellenberg, E. G. (2012). Music cognition: a developmental perspective. *Topics in Cognitive Science*, 4(4):485–497.

Strange, W. (1995). *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research*. York Press, Timonium, MD.

Sussman, E. (2007). A new view on the MMN and attention debate. *Journal of Psychophysiology*, 21(3):164–175.

Tallal, P. and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, 29(7):382–390.

Tremblay, K., Kraus, N., Carrell, T. D., and McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America*, 102(6):3762–3773.

Tremblay, K., Kraus, N., and McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *NeuroReport*, 9(16):3556.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., and Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, 22(2):79–90.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., and Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, 30(33):11177–11187.

Tzovara, A., Rossetti, A. O., Spierer, L., Grivel, J., Murray, M. M., Oddo, M., and De Lucia, M. (2013). Progression of auditory discrimination based on neural decoding predicts awakening from coma. *Brain*, 136(1):81–89.

Valderrama, A. T., Paclik, P., Vansteensel, M. J., Aarnoutse, E. J., and Ramsey, N. F. (2012). Error probability of intracranial brain computer interfaces under non-task elicited brain states. *Clinical Neurophysiology*, 123(12):2392–2401.

van Alphen, P. and Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics*, 32:455–491.

van der Waal, M., Severens, M., Geuze, J., and Desain, P. (2012). Introducing the tactile speller: an ERP-based brain–computer interface for communication. *Journal of Neural Engineering*, 9(4):045002.

van Gerven, M., Bahramisharif, A., Heskes, T., and Jensen, O. (2009). Selecting features for BCI control based on a covert spatial attention paradigm. *Neural Networks*, 22(9):1271–1277.

van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S., and Desain, P. (2009). The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4):041001.

van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., and Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research*, 23(2-3):270–276.

Varma, S., McCandliss, B. D., and Schwartz, D. L. (2008). Scientific and pragmatic challenges for bridging education and neuroscience. *Educational Researcher*, 37(3):140–152.

Vidaurre, C. and Blankertz, B. (2010). Towards a cure for BCI illiteracy. *Brain Topography*, 23(2):194–198.

Vlek, R., Schaefer, R., Gielen, C., Farquhar, J., and Desain, P. (2011). Sequenced subjective accents for brain–computer interfaces. *Journal of Neural Engineering*, 8(3):036002.

Vlek, R. J., Steines, D., Szibbo, D., Kübler, A., Schneider, M. J., Haselager, P., and Nijboer, F. (2012). Ethical issues in brain–computer interface research, development, and dissemination. *Journal of Neurologic Physical Therapy*, 36(2):94–99.

Weiskopf, N. (2012). Real-time fMRI and its application to neurofeedback. *NeuroImage*, 62(2):682–692.

Weisz, N., Hartmann, T., Müller, N., Lorenz, I., and Obleser, J. (2011). Alpha rhythms in audition: Cognitive and clinical perspectives. *Frontiers in Psychology*, 2.

Werker, J. F. and Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, 24(5):672–683.

Werker, J. F. and Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1):49–63.

Winkler, I., Denham, S. L., and Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13(12):532–540.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R. J., and Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36(5):638–642.

Wolpaw, J., Birbaumer, N., and McFarland, D. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, pages 1–25.

Yoo, S., OLeary, H. M., Fairneny, T., Chen, N., Panych, L. P., Park, H., and Jolesz, F. A. (2006). Increasing cortical activity in auditory areas through neurofeedback functional magnetic resonance imaging. *NeuroReport*, 17(12):1273–1278.

Zander, T. O. and Kothe, C. (2011). Towards passive brain–computer interfaces: Applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering*, 8(2):025005.

Zoefel, B., Huster, R. J., and Herrmann, C. S. (2011). Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance. *Neuroimage*, 54(2):1427–1431.

Zotev, V., Krueger, F., Phillips, R., Alvarez, R. P., Simmons, W. K., Bellgowan, P., Drevets, W. C., and Bodurka, J. (2011). Self-regulation of amygdala activation using real-time fMRI neurofeedback. *PLOS ONE*, 6(9):e24522.

Zwiers, M. P., Versnel, H., and Opstal, A. J. V. (2004). Involvement of monkey inferior colliculus in spatial hearing. *Journal of Neuroscience*, 24(17):4145–4156.

# Cirriculum Vitae

Alex Brandmeyer was born on November 24, 1980 in La Jolla, California in the United States. He attended primary and secondary school in Carlsbad, California before going on to pursue bachelor's degrees in English Literature, Philosophy and Cognitive Science at the University of California at Berkeley. After completing these degrees in December 2000, he spent time traveling in Europe before taking a position as a research technician at the Veteran's Affairs Department Speech and Hearing Research Program in Martinez, California under the supervision of Dr. Pierre Divenyi. In 2005, he moved to Nijmegen, the Netherlands to pursue a Master's degree in Cognitive Psychology and to work as a research technician in the Music, Mind, Machine group at the Donders Institute for Brain, Cognition and Behavior under the supervision of Prof. Dr. Peter Desain. In 2009, he was awarded a PhD position, and for the next four years conducted the research that has been presented in this dissertation. Currently, he is working as a postdoctoral researcher at the Max Planck Institute for Human Cognitive and Brain Sciences in the Auditory Cognition group. Alex has been married to Elisenda Peters since 2011, and together they have a beautiful daughter: Alma Gloria Brandmeyer.

# List of Publications

Alex Brandmeyer, Makiko Sadakata, Loukianos Spyrou, James M McQueen & Peter Desain. (2013). Decoding of single-trial auditory mismatch responses for online perceptual monitoring and neurofeedback. *Frontiers in Auditory Cognitive Neuroscience*. 7(265), 1-21.

Alex Brandmeyer, Jason Farquhar, James M McQueen & Peter Desain. (2013). Decoding speech perception by native and non-native speakers using single-trial electrophysiological data. *PLOS ONE*. 8(7), e68261.

Alex Brandmeyer, Peter Desain & James McQueen. (2012). Effects of native language on perceptual sensitivity to phonetic cues. *NeuroReport*. 23(11), 653-657.

Alex Brandmeyer, Renee Timmers, Makiko Sadakata & Peter Desain. (2011). Learning expressive percussion performance under different visual feedback conditions. *Psychological Research*. 75(2), 107-121.

Makiko Sadakata, David Hoppe, Alex Brandmeyer, Renee Timmers & Peter Desain. (2008). Real-Time Visual Feedback for Learning to Perform Short Rhythms with Expressive Variations in Timing and Loudness. *Journal of New Music Research*. 37(3), 207-220.

Pierre Divenyi & Alex Brandmeyer (2005). The role of stress-accent in the understanding of sentences in noise. *Journal of the Acoustical Society of America*. 117(4), 2623.

# Donders Graduate School for Cognitive Neuroscience Series

1. Van Aalderen-Smeets, S.I. (2007). *Neural dynamics of visual selection.* Maastricht University, Maastricht, the Netherlands.

2. Schoffelen, J.M. (2007). *Neuronal communication through coherence in the human motor system.* Radboud University Nijmegen, Nijmegen, the Netherlands.

3. De Lange, F.P. (2008). *Neural mechanisms of motor imagery.* Radboud University Nijmegen, Nijmegen, the Netherlands.

4. Grol, M.J. (2008). *Parieto-frontal circuitry in visuomotor control.* Utrecht University, Utrecht, the Netherlands.

5. Bauer, M. (2008). *Functional roles of rhythmic neuronal activity in the human visual and somatosensory system.* Radboud University Nijmegen, Nijmegen, the Netherlands.

6. Mazaheri, A. (2008). *The influence of ongoing oscillatory brain activity on evoked responses and behaviour.* Radboud University Nijmegen, Nijmegen, the Netherlands.

7. Hooijmans, C.R. (2008). *Impact of nutritional lipids and vascular factors in Alzheimers disease.* Radboud University Nijmegen, Nijmegen, the Netherlands.

8. Gaszner, B. (2008). *Plastic responses to stress by the rodent urocortinergic Edinger-Westphal nucleus.* Radboud University Nijmegen, Nijmegen, the Netherlands.

9. Willems, R.M. (2009). *Neural reflections of meaning in gesture, language and action.* Radboud University Nijmegen, Nijmegen, the Netherlands.

10. Van Pelt, S. (2009). *Dynamic neural representations of human visuomotor space.* Radboud University Nijmegen, Nijmegen, the Netherlands.

11. Lommertzen, J. (2009). *Visuomotor coupling at different levels of complexity.* Radboud University Nijmegen, Nijmegen, the Netherlands.

12. Poljac, E. (2009). *Dynamics of cognitive control in task switching: Looking beyond the switch cost.* Radboud University Nijmegen, Nijmegen, the Netherlands.

13. Poser, B.A. (2009). *Techniques for BOLD and blood volume weighted fMRI.* Radboud University Nijmegen, Nijmegen, the Netherlands.

14. Baggio, G. (2009). *Semantics and the electrophysiology of meaning. Tense, aspect, event structure.* Radboud University Nijmegen, Nijmegen, the Netherlands.

15. Van Wingen, G.A. (2009). *Biological determinants of amygdala functioning.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

16. Bakker, M. (2009). *Supraspinal control of walking: Lessons from motor imagery.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

17. Aarts, E. (2009). *Resisting temptation: The role of the anterior cingulate cortex in adjusting cognitive control.* Radboud University Nijmegen, Nijmegen, the Netherlands.

18. Prinz, S. (2009). *Waterbath stunning of chickens Effects of electrical parameters on the electroencephalogram and physical reflexes of broilers.* Radboud University Nijmegen, Nijmegen, the Netherlands.

19. Knippenberg, J.M.J. (2009). *The N150 of the Auditory Evoked Potential from the rat amygdala: In search for its functional significance.* Radboud University Nijmegen, Nijmegen, the Netherlands.

20. Dumont, G.J.H. (2009). *Cognitive and physiological effects of 3,4-methylenedioxymethamphetamine (MDMA or ecstasy) in combination with alcohol or cannabis in humans.* Radboud University Nijmegen, Nijmegen, the Netherlands.

21. Pijnacker, J. (2010). *Defeasible inference in autism: A behavioral and electrophysiogical approach.* Radboud University Nijmegen, Nijmegen, the Netherlands.

22. De Vrijer, M. (2010). *Multisensory integration in spatial orientation.* Radboud University Nijmegen, Nijmegen, the Netherlands.

23. Vergeer, M. (2010). *Perceptual visibility and appearance: Effects of color and form.* Radboud University Nijmegen, Nijmegen, the Netherlands.

24. Levy, J. (2010). *In cerebro unveiling unconscious mechanisms during reading.* Radboud University Nijmegen, Nijmegen, the Netherlands.

25. Treder, M. S. (2010). *Symmetry in (inter)action.* Radboud University Nijmegen, Nijmegen, the Netherlands.

26. Horlings C.G.C. (2010). *A weak balance: Balance and falls in patients with neuromuscular disorders.* Radboud University Nijmegen, Nijmegen, the Netherlands.

27. Snaphaan, L.J.A.E. (2010). *Epidemiology of post-stroke behavioural consequences.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

28. Dado Van Beek, H.E.A. (2010). *The regulation of cerebral perfusion in patients with Alzheimers disease.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

29. Derks, N.M. (2010). *The role of the non-preganglionic Edinger-Westphal nucleus in sex-dependent stress adaptation in rodents.* Radboud University Nijmegen, Nijmegen, the Netherlands.

30. Wyczesany, M. (2010). *Covariation of mood and brain activity. Integration of subjective self-report data with quantitative EEG measures.* Radboud University Nijmegen, Nijmegen, the Netherlands.

31. Beurze S.M. (2010). *Cortical mechanisms for reach planning.* Radboud University Nijmegen, Nijmegen, the Netherlands.

32. Van Dijk, J.P. (2010). *On the Number of Motor Units.* Radboud University Nijmegen, Nijmegen, the Netherlands.

33. Lapatki, B.G. (2010). *The Facial Musculature - Characterization at a Motor Unit Level.* Radboud University Nijmegen, Nijmegen, the Netherlands.

34. Kok, P. (2010). *Word order and verb inflection in agrammatic sentence production.* Radboud University Nijmegen, Nijmegen, the Netherlands.

35. van Elk, M. (2010). *Action semantics: Functional and neural dynamics.* Radboud University Nijmegen, Nijmegen, the Netherlands.

36. Majdandzic, J. (2010). *Cerebral mechanisms of processing action goals in self and others.* Radboud University Nijmegen, Nijmegen, the Netherlands.

37. Snijders, T.M. (2010). *More than words - Neural and genetic dynamics of syntactic unification.* Radboud University Nijmegen, Nijmegen, the Netherlands.

38. Grootens, K.P. (2010). *Cognitive dysfunction and effects of antipsychotics in schizophrenia and borderline personality disorder.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

39. Nieuwenhuis, I.L.C. (2010). *Memory consolidation: A process of integration Converging evidence from MEG, fMRI and behavior.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

40. Menenti, L.M.E. (2010). *The right language: Differential hemispheric contributions to language production and comprehension in context.* Radboud University Nijmegen, Nijmegen, the Netherlands.

41. Van Dijk, H.P. (2010). *The state of the brain, how alpha oscillations shape behaviour and event related responses.* Radboud University Nijmegen, Nijmegen, the Netherlands.

42. Meulenbroek, O.V. (2010). *Neural correlates of episodic memory in healthy aging and Alzheimers disease.* Radboud University Nijmegen, Nijmegen, the Netherlands.

43. Oude Nijhuis, L.B. (2010). *Modulation of human balance reactions.* Radboud University Nijmegen, Nijmegen, the Netherlands.

44. Qin, S. (2010). *Adaptive memory: Imaging medial temporal and prefrontal memory systems.* Radboud University Nijmegen, Nijmegen, the Netherlands.

45. Timmer, N.M. (2011). *The interaction of heparan sulfate proteoglycans with the amyloid protein.* Radboud University Nijmegen, Nijmegen, the Netherlands.

46. Crajé, C. (2011). *(A)typical motor planning and motor imagery.* Radboud University Nijmegen, Nijmegen, the Netherlands.

47. Van Grootel, T.J. (2011). *On the role of eye and head position in spatial localisation behaviour.* Radboud University Nijmegen, Nijmegen, the Netherlands.

48. Lamers, M.J.M. (2011). *Levels of selective attention in action planning.* Radboud University Nijmegen, Nijmegen, the Netherlands.

49. Van der Werf, J. (2011). *Cortical oscillatory activity in human visuomotor integration.* Radboud University Nijmegen, Nijmegen, the Netherlands.

50. Scheeringa, R. (2011). *On the relation between oscillatory EEG activity and the BOLD signal.* Radboud University Nijmegen, Nijmegen, the Netherlands.

51. Bögels, S. (2011). *The role of prosody in language comprehension: When prosodic breaks and pitch accents come into play.* Radboud University Nijmegen, Nijmegen, the Netherlands.

52. Ossewaarde, L. (2011). *The mood cycle: Hormonal influences on the female brain.* Radboud University Nijmegen, Nijmegen, the Netherlands.

53. Kuribara, M. (2011). *Environment-induced activation and growth of pituitary melanotrope cells of Xenopus laevis.* Radboud University Nijmegen, Nijmegen, the Netherlands.

54. Helmich, R.C.G. (2011). *Cerebral reorganization in Parkinsons disease.* Radboud University Nijmegen, Nijmegen, the Netherlands.

55. Boelen, D. (2011). *Order out of chaos? Assessment and treatment of executive disorders in brain-injured patients.* Radboud University Nijmegen, Nijmegen, the Netherlands.

56. Koopmans, P.J. (2011). *fMRI of cortical layers.* Radboud University Nijmegen, Nijmegen, the Netherlands.

57. van der Linden, M.H. (2011). *Experience-based cortical plasticity in object category representation.* Radboud University Nijmegen, Nijmegen, the Netherlands.

58. Kleine, B.U. (2011). *Motor unit discharges - Physiological and diagnostic studies in ALS.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

59. Paulus, M. (2011). *Development of action perception: Neurocognitive mechanisms underlying childrens processing of others actions.* Radboud University Nijmegen, Nijmegen, the Netherlands.

60. Tieleman, A.A. (2011). *Myotonic dystrophy type 2. A newly diagnosed disease in the Netherlands.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

61. Van Leeuwen, T.M. (2011). *How one can see what is not there: Neural mechanisms of grapheme-colour synaesthesia.* Radboud University Nijmegen, Nijmegen, the Netherlands.

62. Van Tilborg, I.A.D.A. (2011). *Procedural learning in cognitively impaired patients and its application in clinical practice.* Radboud University Nijmegen, Nijmegen, the Netherlands.

63. Bruinsma, I.B. (2011). *Amyloidogenic proteins in Alzheimers disease and Parkinsons disease: Interaction with chaperones and inflammation.* Radboud University Nijmegen, Nijmegen, the Netherlands.

64. Voermans, N. (2011). *Neuromuscular features of Ehlers-Danlos syndrome and Marfan syndrome; expanding the phenotype of inherited connective tissue disorders and investigating the role of the extracellular matrix in muscle.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

65. Reelick, M. (2011). *One step at a time. Disentangling the complexity of preventing falls in frail older persons.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

66. Buur, P.F. (2011). *Imaging in motion. Applications of multi-echo fMRI.* Radboud University Nijmegen, Nijmegen, the Netherlands.

67. Schaefer, R.S. (2011). *Measuring the minds ear: EEG of music imagery.* Radboud University Nijmegen, Nijmegen, the Netherlands.

68. Xu, L. (2011). *The non-preganglionic Edinger-Westphal nucleus: An integration center for energy balance and stress adaptation.* Radboud University Nijmegen, Nijmegen, the Netherlands.

69. Schellekens, A.F.A. (2011). *Gene-environment interaction and intermediate phenotypes in alcohol dependence.* Radboud University Nijmegen, Nijmegen, the Netherlands.

70. Van Marle, H.J.F. (2011). *The amygdala on alert: A neuroimaging investigation into amygdala function during acute stress and its aftermath.* Radboud University Nijmegen, Nijmegen, the Netherlands.

71. De Laat, K.F. (2011). *Motor performance in individuals with cerebral small vessel disease: An MRI study.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

72. Mädebach, A. (2011). *Lexical access in speaking: Studies on lexical selection and cascading activation.* Radboud University Nijmegen, Nijmegen, the Netherlands.

73. Poelmans, G.J.V. (2011). *Genes and protein networks for neurodevelopmental disorders.* Radboud University Nijmegen, Nijmegen, the Netherlands.

74. Van Norden, A.G.W. (2011). *Cognitive function in elderly individuals with cerebral small vessel disease.* An MRI study. Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

75. Jansen, E.J.R. (2011). *New insights into V-ATPase functioning: the role of its accessory subunit Ac45 and a novel brain-specific Ac45 paralog.* Radboud University Nijmegen, Nijmegen, the Netherlands.

76. Haaxma, C.A. (2011). *New perspectives on preclinical and early stage Parkinson's disease.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

77. Haegens, S. (2012). *On the functional role of oscillatory neuronal activity in the somatosensory system.* Radboud University Nijmegen, Nijmegen, the Netherlands.

78. van Barneveld, D.C.P.B.M. (2012). *Integration of exteroceptive and interoceptive cues in spatial localization.* Radboud University Nijmegen, Nijmegen, the Netherlands.

79. Spies, P.E. (2012). *The reflection of Alzheimer disease in CSF.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

80. Helle, M. (2012). *Artery-specific perfusion measurements in the cerebral vasculature by magnetic resonance imaging.* Radboud University Nijmegen, Nijmegen, the Netherlands.

81. Egetemeir, J. (2012). *Neural correlates of real-life joint action.* Radboud University Nijmegen, Nijmegen, the Netherlands.

82. Janssen, L. (2012). *Planning and execution of (bi)manual grasping.* Radboud University Nijmegen, Nijmegen, the Netherlands.

83. Vermeer, S. (2012). *Clinical and genetic characterisation of autosomal recessive cerebellar ataxias.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

84. Vrins, S. (2012). *Shaping object boundaries: Contextual effects in infants and adults.* Radboud University Nijmegen, Nijmegen, the Netherlands.

85. Weber, K.M. (2012). *The language learning brain: Evidence from second language and bilingual studies of syntactic processing.* Radboud University Nijmegen, Nijmegen, the Netherlands.

86. Verhagen, L. (2012). *How to grasp a ripe tomato.* Utrecht University, Utrecht, the Netherlands.

87. Nonkes, L.J.P. (2012). *Serotonin transporter gene variance causes individual differences in rat behaviour: For better and for worse.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

88. Joosten-Weyn Banningh, L.W.A. (2012). *Learning to live with Mild Cognitive Impairment: development and evaluation of a psychological intervention for patients with Mild Cognitive Impairment and their significant others.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

89. Xiang, HD. (2012). *The language networks of the brain.* Radboud University Nijmegen, Nijmegen, the Netherlands.

90. Snijders, A.H. (2012). *Tackling freezing of gait in Parkinson's disease.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

91. Rouwette, T.P.H. (2012). *Neuropathic pain and the brain - Differential involvement of corticotropin-releasing factor and urocortin 1 in acute and chronic pain processing.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

92. Van de Meerendonk, N. (2012). *States of indecision in the brain: Electrophysiological and hemodynamic reflections of monitoring in visual language perception.* Radboud University Nijmegen, Nijmegen, the Netherlands.

93. Sterrenburg, A. (2012). *The stress response of forebrain and midbrain regions: Neuropeptides, sex-specificity and epigenetics.* Radboud University Nijmegen, Nijmegen, The Netherlands.

94. Uithol, S. (2012). *Representing action and intention.* Radboud University Nijmegen, Nijmegen, The Netherlands.

95. Van Dam, W.O. (2012). *On the specificity and flexibility of embodied lexical-semantic representations.* Radboud University Nijmegen, Nijmegen, The Netherlands.

96. Slats, D. (2012). *CSF biomarkers of Alzheimers disease: Serial sampling analysis and the study of circadian rhythmicity.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

97. Van Nuenen, B.F.L. (2012). *Cerebral reorganization in premotor parkinsonism.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

98. van Schouwenburg, M.R. (2012). *Fronto-striatal mechanisms of attentional control.* Radboud University Nijmegen, Nijmegen, The Netherlands.

99. Azar, M.G. (2012). *On the theory of reinforcement learning: Methods, convergence analysis and sample complexity.* Radboud University Nijmegen, Nijmegen, The Netherlands.

100. Meeuwissen, E.B. (2012). *Cortical oscillatory activity during memory formation.* Radboud University Nijmegen, Nijmegen, The Netherlands.

101. Arnold, J.F. (2012). *When mood meets memory: Neural and behavioral perspectives on emotional memory in health and depression.* Radboud University Nijmegen, Nijmegen, The Netherlands.

102. Gons, R.A.R. (2012). *Vascular risk factors in cerebral small vessel disease: A diffusion tensor imaging study.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

103. Wingbermühle, E. (2012). *Cognition and emotion in adults with Noonan syndrome: A neuropsychological perspective.* Radboud University Nijmegen, Nijmegen, The Netherlands.

104. Walentowska, W. (2012). *Facing emotional faces. The nature of automaticity of facial emotion processing studied with ERPs.* Radboud University Nijmegen, Nijmegen, The Netherlands.

105. Hoogman, M. (2012). *Imaging the effects of ADHD risk genes.* Radboud University Nijmegen, Nijmegen, The Netherlands.

106. Tramper, J. J. (2012). *Feedforward and feedback mechanisms in sensory motor control.* Radboud University Nijmegen, Nijmegen, The Netherlands.

107. Van Eijndhoven, P. (2012). *State and trait characteristics of early course major depressive disorder.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

108. Visser, E. (2012). *Leaves and forests: Low level sound processing and methods for the large-scale analysis of white matter structure in autism.* Radboud University Nijmegen, Nijmegen, The Netherlands.

109. Van Tooren-Hoogenboom, N. (2012). *Neuronal communication in the synchronized brain. Investigating the functional role of visually-induced gamma band activity: Lessons from MEG.* Radboud University Nijmegen, Nijmegen, The Netherlands.

110. Henckens, M.J.A.G. (2012). *Imaging the stressed brain. Elucidating the time- and region-specific effects of stress hormones on brain function: A translational approach.* Radboud University Nijmegen, Nijmegen, The Netherlands.

111. Van Kesteren, M.T.R. (2012). *Schemas in the brain: Influences of prior knowledge on learning, memory, and education.* Radboud University Nijmegen, Nijmegen, The Netherlands.

112. Brenders, P. (2012). *Cross-language interactions in beginning second language learners.* Radboud University Nijmegen, Nijmegen, The Netherlands.

113. Ter Horst, A.C. (2012). *Modulating motor imagery. Contextual, spatial and kinaesthetic influences.* Radboud University Nijmegen, Nijmegen, The Netherlands.

114. Tesink, C.M.J.Y. (2013). *Neurobiological insights into language comprehension in autism: Context matters.* Radboud University Nijmegen, Nijmegen, The Netherlands.

115. Böckler, A. (2013). *Looking at the world together. How others attentional relations to jointly attended scenes shape cognitive processing.* Radboud University Nijmegen, Nijmegen, The Netherlands.

116. Van Dongen, E.V. (2013). *Sleeping to Remember. On the neural and behavioral mechanisms of sleep-dependent memory consolidation.* Radboud University Nijmegen, Nijmegen, The Netherlands.

117. Volman, I. (2013). *The neural and endocrine regulation of emotional actions.* Radboud University Nijmegen, Nijmegen, The Netherlands.

118. Buchholz, V. (2013). *Oscillatory activity in tactile remapping.* Radboud University Nijmegen, Nijmegen, The Netherlands.

119. Van Deurzen, P.A.M. (2013). *Information processing and depressive symptoms in healthy adolescents.* Radboud University Nijmegen, Nijmegen, The Netherlands.

120. Whitmarsh, S. (2013). *Nonreactivity and metacognition in mindfulness.* Radboud University Nijmegen, Nijmegen, The Netherlands.

121. Vesper, C. (2013). *Acting together: Mechanisms of intentional coordination.* Radboud University Nijmegen, Nijmegen, The Netherlands.

122. Lagro, J. (2013). *Cardiovascular and cerebrovascular physiological measurements in clinical practice and prognostics in geriatric patients.* Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

123. Eskenazi, T.T. (2013). *You, us & them: From motor simulation to ascribed shared intentionality in social perception.* Radboud University Nijmegen, Nijmegen, The Netherlands.

124. Ondobaka, S. (2013). *On the conceptual and perceptual processing of own and others behavior.* Radboud University Nijmegen, Nijmegen, The Netherlands.

125. Overvelde, J.A.A.M. (2013). *Which practice makes perfect? Experimental studies on the acquisition of movement sequences to identify the best learning condition in good and poor writers.* Radboud University Nijmegen, Nijmegen, The Netherlands.

126. Kalisvaart, J.P. (2013). *Visual ambiguity in perception and action.* Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

127. Kroes, M. (2013). *Altering memories for emotional experiences.* Radboud University Nijmegen, Nijmegen, The Netherlands.

128. Duijnhouwer, J. (2013). *Studies on the rotation problem in self-motion perception.* Radboud University Nijmegen, Nijmegen, The Netherlands.

129. Nijhuis, E.H.J (2013). *Macroscopic networks in the human brain: Mapping connectivity in healthy and damaged brains.* University of Twente, Enschede, The Netherlands.

130. Braakman, M. H. (2013). *Posttraumatic stress disorder with secondary psychotic features. A diagnostic validity study among refugees in the Netherlands.* Radboud University Nijmegen, Nijmegen, The Netherlands.

131. Zedlitz, A.M.E.E. (2013). *Brittle brain power. Post-stroke fatigue, explorations into assessment and treatment.* Radboud University Nijmegen, Nijmegen, The Netherlands.

132. Schoon, Y. (2013). *From a gait and falls clinic visit towards self-management of falls in frail elderly.* Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

133. Jansen, D. (2013). *The role of nutrition in Alzheimer's disease - A study in transgenic mouse models for Alzheimer's disease and vascular disorders.* Radboud University Nijmegen, Nijmegen, The Netherlands.

134. Kos, M. (2013). *On the waves of language - Electrophysiological reflections on semantic and syntactic processing.* Radboud University Nijmegen, Nijmegen, The Netherlands.

135. Severens, M. (2013). *Towards clinical BCI applications: Assistive technology and gait rehabilitation.* Radboud University Nijmegen, Nijmegen, Sint Maartenskliniek, Nijmegen, The Netherlands.

136. Bergmann, H. (2014). *Two is not always better than one: On the functional and neural (in)dependence of working memory and long-term memory.* Radboud University Nijmegen, Nijmegen, The Netherlands.

137. Wronka, E. (2013). *Searching for the biological basis of human mental abilitites. The relationship between attention and intelligence studied with P3.* Radboud University Nijmegen, Nijmegen, The Netherlands.

138. Lüttjohann, A.K. (2013). *The role of the cortico-thalamo-cortical system in absence epilepsy.* Radboud University Nijmegen, Nijmegen, The Netherlands.

139. Brazil, I.A. (2013). *Change doesnt come easy: Dynamics of adaptive behavior in psychopathy.* Radboud University Nijmegen, Nijmegen, The Netherlands.

140. Zerbi, V. (2013). *Impact of nutrition on brain structure and function. A magnetic resonance imaging approach in Alzheimer mouse models.* Radboud University Nijmegen, Nijmegen, The Netherlands.

141. Delnooz, C.C.S. (2014). *Unravelling primary focal dystonia. A treatment update and new pathophysiological insights.* Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

142. Bultena, S.S. (2013). *Bilingual processing of cognates and language switches in sentence context.* Radboud University Nijmegen, Nijmegen, The Netherlands.

143. Janssen, G. (2014). *Diagnostic assessment of psychiatric patients: A contextual perspective on executive functioning.* Radboud University Nijmegen, Nijmegen, The Netherlands.

144. Piai, V. Magalhães (2014). *Choosing our words: Lexical competition and the involvement of attention in spoken word production.* Radboud University Nijmegen, Nijmegen, The Netherlands.

145. Van Ede, F. (2014). *Preparing for perception. On the attentional modulation, perceptual relevance and physiology of oscillatory neural activity.* Radboud University Nijmegen, Nijmegen, The Netherlands.

146. Brandmeyer, A. (2014). *Auditory perceptual learning via decoded EEG neurofeedback: a novel paradigm.* Radboud University Nijmegen, Nijmegen, The Netherlands.